e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR TECHNOLOGY, UIS
https://ejpi.uis.edu.my/index.php/ejpi
e-ISSN 2289-7909

VOL. 13 NO. 1 (APRIL, 2026): page 64-88
DOI: 10.53840/e-jpi.v13i1.348

Received: January 20, 2026 | Accepted: April 28, 2026 | Published: April 30, 2026

Assessing Literacy Competencies Among Polytechnic Lecturers in Malaysia Using a Rasch-Based Measurement Model

Normazita Binti Mat Ali[1], Nurfirdawati Binti Muhamad Hanafi[2], Razif Salim[3], Sani Inusa Milala[4]

[1,2]Faculty of Education, Universiti Tun Hussein Onn Malaysia 86400 Parit Raja, Batu Pahat, johor Malaysia
[1]Tun Syed Nasir Syed Ismail Polytechnic Hub Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, Malaysia
[3]Tun amina library, Universiti Tun Hussein Onn Malaysia 86400 Parit Raja, Batu Pahat, johor Malaysia
[4]Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia 86400 Parit Raja, Batu Pahat, johor Malaysia.
Email: [1]normazita81@gmail.com, normazitamatali@gmail.com, [2]nurfirda@uthm.edu.my, [3]razifs@uthm.edu.my, [4]sinusamilala@gmail.com

ABSTRACT

Assessment literacy is a critical component of effective teaching and learning within Outcome-Based Education (OBE) systems, particularly in polytechnic institutions where competency-based outcomes are emphasized. Despite ongoing assessment reforms in Malaysian polytechnics, concerns remain regarding lecturers' literacy competencies, especially in aligning assessment practices with OBE principles. Conventional evaluation methods often fail to capture latent competency levels and item functioning, limiting the accuracy of assessment literacy measurement. There is a need for a robust and psychometrically sound measurement approach that can accurately assess lecturers' literacy competencies and reveal underlying cognitive demands associated with assessment practices. Employing Rasch measurement addresses this gap by providing interval-level estimates, enabling more precise evaluation of lecturer competencies and supporting evidence-based professional development and policy decisions in Malaysian polytechnic education. This study adopted a quantitative research design using a Rasch-based measurement model to assess literacy competencies among polytechnic lecturers in Malaysia. A structured assessment instrument focusing on Outcome-Based Education and assessment literacy was administered to a sample of 33 polytechnic lecturers. Rasch analysis was conducted to estimate item difficulty, evaluate item targeting, and examine construct representation. Descriptive statistics and Rasch logit measures were used to analyze item functioning and respondent performance. The Rasch analysis revealed a well-ordered item difficulty hierarchy ranging from −0.835 to +1.987 logits, indicating strong discrimination across varying competency levels. Items assessing foundational assessment knowledge were identified as easy, while those related to formative assessment and core OBE concepts were significantly more difficult. Although some items recorded high correct response rates (up to 87.9%), Rasch logit estimates highlighted latent conceptual complexity not captured by classical statistics alone. The scale demonstrated effective item targeting, minimal item clustering, and adequate construct coverage, supporting its measurement validity. The findings confirm the effectiveness of Rasch modeling in assessing literacy competencies among polytechnic lecturers and reveal critical gaps in advanced OBE-related assessment literacy. The study underscores the importance of using robust measurement models for evaluating educational competencies and provides empirical evidence to inform targeted professional development and assessment policy enhancement in Malaysian polytechnic institution.

Keywords: Assessment Literacy; Outcome-Based Education; Rasch Measurement Model; Polytechnic Lecturers; Malaysia

INTRODUCTION

The global economy increasingly demands a highly skilled workforce, positioning Technical and Vocational Education and Training (TVET) institutions as central to meeting this need (Gyimah, 2020; Komaro et al., 2022). In Malaysia, polytechnics play a pivotal role in producing job-ready graduates aligned with national development plans and the Sustainable Development Goals (Nalathambi et al., 2023).

National policies such as the Malaysia Education Blueprint 2015–2025 emphasize Outcome-Based Education (OBE), shifting focus from teaching inputs to measurable learning outcomes (Suo, 2023; Rao, 2020). Within this framework, assessment becomes a continuous process that evaluates cognitive, psychomotor, and affective competencies (Tenedero & Pacadaljen, 2021; Agi et al., 2018).

Polytechnic lecturers are therefore expected to be assessment-literate, capable of designing valid and reliable evaluations aligned with Program and Course Learning Outcomes (Naqvi et al., 2019; Popham, 2018). Yet, a gap persists between policy expectations and lecturers' practical assessment literacy, due to limited professional development and administrative burdens (Asim et al., 2021; Herd & Moynihan, 2019).

Existing instruments often fail to capture the unique demands of Malaysia's TVET context, particularly in assessing psychomotor and industry-related skills (Mustapha & Hussain, 2022; Komaro et al., 2022). To address this, the present study employs the Rasch measurement model to develop and validate a context-specific tool for assessing literacy competencies among polytechnic lecturers (Setiawan et al., 2025)

The Rasch model provides a robust psychometric framework that transforms raw ordinal data into interval-level measurements, enabling a rigorous evaluation of item difficulty, person ability, and the unidimensionality of the construct (Bond & Fox, 2015; Komaro et al., 2022; Setiawan et al., 2025). The development of this validated instrument will provide a critical tool for polytechnic administrators and educational researchers to accurately gauge assessment literacy, thereby informing targeted professional development and ultimately enhancing the quality of teaching and learning within Malaysia's vital TVET sector (Mustapha & Hussain, 2022; Lérias et al., 2024).

Globally, the critical role of Technical and Vocational Education and Training (TVET) in driving economic development and supplying a skilled workforce is well-established, creating an urgent need for robust quality assurance mechanisms within these institutions (Gyimah, 2020; Yusof et al., 2023). In the Malaysian context, this is reflected in ambitious national reforms, such as those outlined in the Malaysia Education Blueprint 2015–2025, which mandate a systemic shift towards Outcome-Based Education (OBE) to enhance graduate quality and international competitiveness (Suo, 2023; Rao, 2020; Ling, 2023). Consequently, Malaysian polytechnics are under significant pressure to realign their curricula and, most critically, their assessment practices to accurately measure complex learning outcomes across cognitive, psychomotor, and affective domains, thereby ensuring graduates meet industry standards (Adnan, Paimin, & Hasan, 2019; Nalathambi et al., 2023; Lukman & Victor, 2024).

The successful implementation of this OBE paradigm is entirely dependent on lecturers possessing a high degree of assessment literacy the ability to design, implement, and interpret effective assessments that inform teaching and improve learning (Popham, 2018; Suherman, 2022). However, a significant problem exists as studies indicate a pervasive gap between OBE policy expectations and actual classroom assessment practices, with educators often struggling to translate theoretical frameworks into valid and reliable evaluation methods, particularly for performance-based and affective learning outcomes (Asim et al., 2021; Agi, Aduloju, & Kpum, 2018; Hambali et al., 2024).

PROBLEM STATEMENT

This gap is acutely felt in the polytechnic environment, where the pressure to align with industry needs and comply with rigorous academic audits, such as the Updated Academic Performance Audit, further compounds the challenge (Razalli et al., 2023). While previous research has extensively documented the importance of outcome-based education (OBE) and curriculum design in technical and vocational education and training (TVET) (Azmi & Salleh, 2021; Mustapha & Hussain, 2022), and other studies have explored general assessment literacy (Koh et al., 2018; Levi & Inbar-Lourie, 2020), a critical void remains. Specifically, existing studies fail to provide a psychometrically robust and context-specific instrument to diagnose the multifaceted assessment literacy required by polytechnic lecturers within the unique demands of the Malaysian OBE and TVET ecosystem. For instance, Marzuki et al. (2022) focus on financial allocation, while Adnan et al. (2019) concentrate on curriculum readiness for industry, but neither addresses the core competency of assessment (Marzuki et al., 2022; Adnan et al., 2019). Similarly, Komaro et al. (2022) highlight vocational lecturers' pedagogical competence needs in Indonesia, yet their work does not extend to the measurement of assessment literacy in Malaysian polytechnics (Komaro et al., 2022). Other studies have examined literacy in different contexts, such as artificial intelligence literacy in higher education (Lérias et al., 2024) and language assessment literacy among EFL teachers in Indonesia (Suherman, 2022), but these remain contextually distant from the Malaysian TVET system. Ling (2023) investigates gender influences on information literacy among Malaysian polytechnic students, yet this focuses on learner competencies rather than lecturer assessment literacy. Likewise, Setiawan et al. (2025) assess pedagogical competence among vocational secondary school teachers, which, while relevant to teaching quality, does not provide a validated tool for measuring assessment literacy in higher-level TVET institutions (Setiawan et al., 2025). Research in other regions also highlights related but incomplete perspectives. Lukman and Victor (2024) evaluate lecturers' professional competency in student assessment in Nigeria, and Al-Bahlani and Ecke (2023) explore digital assessment literacy among postsecondary English language teachers in Oman. Hambali et al. (2024) investigate language assessment literacy in EFL edupreneurship, while Yusof et al. (2023) identify competencies needed for improving teaching among Nigerian polytechnic lecturers. Although these studies underscore the importance of assessment competence, they remain either geographically or contextually misaligned with the Malaysian OBE-driven TVET ecosystem (Lukman & Victor, 2024; Al-Bahlani & Ecke, 2023; Hambali et al., 2024; Yusof et al., 2023). Therefore, despite the breadth of existing literature, a significant gap persists: no study has developed a psychometrically validated, context-specific instrument capable of diagnosing the complex assessment literacy competencies required of Malaysian polytechnic lecturers within the OBE-driven TVET framework. This missing link continues to hinder targeted professional development and risks misalignment between policy expectations and classroom realities. To address this gap, the present study employs the Rasch measurement model to construct and validate a precise instrument that empirically gauges assessment literacy levels among polytechnic lecturers, thereby offering a foundational tool for faculty development and enhancing the quality of TVET delivery in Malaysia.

Novelty

While prior studies have examined aspects of teaching and assessment, none provide a psychometrically validated, context-specific tool for measuring assessment literacy among Malaysian polytechnic lecturers. Komaro et al. (2022) and Setiawan et al. (2025) focused on pedagogical competence, while Ling (2023) and Lérias et al. (2024) addressed student or AI literacy, leaving lecturer assessment skills unmeasured. Suherman (2022), Hambali et al. (2024), and Al-Bahlani and Ecke (2023) explored language or digital assessment literacy in other contexts, and Lukman and Victor (2024) along with Yusof et al. (2023) studied lecturer competencies in Nigeria, but none aligned with Malaysia's OBE-driven TVET system. Collectively, these works highlight the importance of assessment competence yet fail to address the unique demands of Malaysian polytechnics. The novelty of this study lies in employing the Rasch model to construct and validate a rigorous, context-specific instrument that directly measures assessment literacy, filling a critical methodological and practical gap for targeted faculty development and improved TVET delivery.

LITERATURE REVIEW

Literacy Assessment

Assessment literacy, defined as educators' capacity to understand, design, implement, and utilize assessment practices effectively (Popham, 2018; Suherman, 2022), serves as the core theoretical construct of this study. In the Malaysian polytechnic context, this construct extends beyond general assessment knowledge to incorporate three critical dimensions aligned with Outcome-Based Education (OBE) and Technical and Vocational Education and Training (TVET) demands (Mustapha & Hussain, 2022; Ling, 2023). *First*, the technical dimension involves designing authentic, performance-based assessments that measure psychomotor skills and practical competencies, mirroring real-world industry requirements (Pratiwi, Arief, & Churiyah, 2018; Komaro et al., 2022; Adnan, Paimin, & Hasan, 2019; Yusof et al., 2023). *Second*, the pedagogical dimension requires integrating assessment with teaching and learning by aligning tasks with Course and Program Learning Outcomes (CLOs/PLOs) and employing diverse methods such as self- and peer assessment, using assessment to improve instruction and student engagement (Seifert & Feliks, 2019; Meijer et al., 2020; Setiawan et al., 2025; Lukman & Victor, 2024). *Third*, the socio-contextual dimension addresses the capacity to navigate industry alignment and national quality assurance frameworks, including the Updated Academic Performance Audit standards, ensuring compliance with accreditation while maintaining relevance to industry expectations (Razalli et al., 2023; Lérias et al., 2024; Hambali et al., 2024; Al-Bahlani & Ecke, 2023)

Components of Literacy Assessment

Assessment literacy is a multidimensional construct comprising three interrelated components (Popham, 2018; Suherman, 2022). *First*, the technical dimension refers to educators' ability to design valid, reliable, and authentic assessment tasks that measure cognitive knowledge alongside psychomotor and practical skills, emphasizing performance-based assessments that replicate real-world industry requirements (Pratiwi, Arief, & Churiyah, 2018; Komaro et al., 2022; Adnan, Paimin, & Hasan, 2019; Yusof et al., 2023). *Second*, the pedagogical dimension highlights the integration of assessment into teaching and learning, requiring lecturers to align tasks with Course and Program Learning Outcomes (CLOs/PLOs) and employ formative feedback, peer assessment, and self-assessment to enhance student engagement, treating assessment as a continuous instructional tool rather than a terminal grading activity (Seifert & Feliks, 2019; Meijer et al., 2020; Setiawan et al., 2025; Lukman & Victor, 2024; Popham, 2018; Suherman, 2022). *Third*, the socio-contextual dimension involves situating assessment practices within institutional, policy, and industry frameworks, including compliance with national quality assurance mechanisms such as the Updated Academic Performance Audit and responsiveness to accreditation requirements, ensuring relevance and credibility (Razalli et al., 2023; Lérias et al., 2024; Hambali et al., 2024; Al-Bahlani & Ecke, 2023; Ling, 2023; Yusof et al., 2023).

Factors Affecting Literacy Assessment

Literacy assessment is shaped by a complex interplay of individual, institutional, and contextual factors that determine how effectively educators design, implement, and interpret assessment practices (Popham, 2018; Suherman, 2022).

i. Teacher Knowledge and Skills

The foundation of assessment literacy lies in educators' understanding of core principles such as validity, reliability, fairness, and alignment with learning outcomes (Popham, 2018; Koh et al., 2018). Teachers with limited psychometric knowledge often struggle to design assessments that accurately measure student competencies, particularly in technical and vocational contexts (Suherman, 2022; Hambali et al., 2024). This lack of expertise can result in assessments that fail to capture the full range of cognitive, psychomotor, and affective skills required in TVET (Komaro et al., 2022; Yusof et al., 2023). The foundation of assessment literacy lies in educators' understanding of core principles such as validity, reliability, fairness, and alignment with learning outcomes (Popham, 2018; Koh et al., 2018). These principles provide the theoretical and practical basis for designing assessments that are

both accurate and meaningful, ensuring that student performance is measured in ways that reflect actual learning achievements (Suherman, 2022; Al-Bahlani & Ecke, 2023). Without a strong grasp of these fundamentals, lecturers risk implementing assessment practices that are inconsistent, subjective, or misaligned with curricular objectives (Hambali et al., 2024; Ling, 2023).

Teachers with limited psychometric knowledge often struggle to design assessments that accurately measure student competencies, particularly in technical and vocational contexts where psychomotor and applied skills are central (Suherman, 2022; Hambali et al., 2024). In such environments, lecturers must go beyond traditional paper-based tests to incorporate authentic, performance-based assessments that mirror industry requirements (Komaro et al., 2022; Adnan, Paimin, & Hasan, 2019). However, the absence of specialized training in psychometric design and measurement frequently results in assessments that fail to capture the complexity of student learning outcomes (Setiawan et al., 2025; Lukman & Victor, 2024).

This lack of expertise can result in assessments that fail to capture the full range of cognitive, psychomotor, and affective skills required in TVET (Komaro et al., 2022; Yusof et al., 2023). For example, while cognitive knowledge may be relatively straightforward to evaluate through written examinations, psychomotor skills such as laboratory performance or technical task execution demand more sophisticated assessment instruments (Razalli et al., 2023; Mustapha & Hussain, 2022). Similarly, affective domains such as professionalism, teamwork, and ethical responsibility are often overlooked or inadequately measured due to lecturers' limited familiarity with holistic assessment frameworks (Lérias et al., 2024; Hambali et al., 2024).

ii. Professional Development and Training

Continuous professional development is a critical factor influencing assessment literacy (Davison & Michell, 2014; Al-Bahlani & Ecke, 2023). Structured training programs, workshops, and mentoring initiatives enhance lecturers' ability to apply diverse assessment methods effectively (Setiawan et al., 2025; Lukman & Victor, 2024). However, many educators report insufficient opportunities for targeted training, leaving gaps in their ability to translate theoretical frameworks into classroom practice (Mustapha & Hussain, 2022; Ling, 2023). Continuous professional development is a critical factor influencing assessment literacy, as it equips educators with the necessary skills to design, implement, and interpret effective assessments (Davison & Michell, 2014; Al-Bahlani & Ecke, 2023). Professional development initiatives provide lecturers with updated knowledge of assessment principles, psychometric techniques, and innovative practices that align with Outcome-Based Education (OBE) and Technical and Vocational Education and Training (TVET) requirements (Komaro et al., 2022; Suherman, 2022). Without such training, lecturers often rely on outdated or limited approaches, which can undermine the validity and reliability of assessment practices (Hambali et al., 2024; Ling, 2023).

Structured training programs, workshops, and mentoring initiatives enhance lecturers' ability to apply diverse assessment methods effectively (Setiawan et al., 2025; Lukman & Victor, 2024). These programs expose educators to authentic, performance-based assessment strategies that capture cognitive, psychomotor, and affective domains, thereby ensuring alignment with industry expectations (Adnan, Paimin, & Hasan, 2019; Yusof et al., 2023). Mentoring and peer-learning opportunities also foster collaborative knowledge sharing, enabling lecturers to refine their practices through feedback and collective problem-solving (Razalli et al., 2023; Lérias et al., 2024).

However, many educators report insufficient opportunities for targeted training, leaving gaps in their ability to translate theoretical frameworks into classroom practice (Mustapha & Hussain, 2022; Ling, 2023). In Malaysia, while national policies emphasize OBE and quality assurance, institutional support for professional development is often inconsistent, resulting in uneven levels of assessment literacy among lecturers (Azmi & Salleh, 2021; Popham, 2018). This lack of systematic training contributes to difficulties in designing assessments that measure complex skills and in meeting accreditation standards (Razalli et al., 2023; Lukman & Victor, 2024).

Ultimately, professional development and training serve as the bridge between policy expectations and classroom realities, ensuring that lecturers possess the competencies required to implement effective assessment practices (Davison & Michell, 2014; Al-Bahlani & Ecke, 2023). Strengthening these initiatives is essential for Malaysian polytechnics to achieve the goals of OBE and to produce graduates who are both academically competent and industry-ready (Komaro et al., 2022; Yusof et al., 2023)

iii. Institutional and Policy Frameworks

National education policies and institutional guidelines strongly shape assessment literacy (Razalli et al., 2023; Azmi & Salleh, 2021). In Malaysia, the Malaysia Education Blueprint 2015–2025 and OBE reforms require lecturers to align assessments with CLOs and PLOs, but implementation challenges persist (Suo, 2023; Rao, 2020). Institutional audits and accreditation requirements further pressure lecturers to comply with quality assurance standards, often without adequate support mechanisms (Mustapha & Hussain, 2022; Komaro et al., 2022). National education policies and institutional guidelines strongly shape assessment literacy, as they establish the standards and expectations that lecturers must meet in their teaching and evaluation practices (Razalli et al., 2023; Azmi & Salleh, 2021). In Malaysia, the Malaysia Education Blueprint 2015–2025 and Outcome-Based Education (OBE) reforms require lecturers to align assessments with Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs), ensuring that student achievement is measured against clearly defined competencies (Suo, 2023; Rao, 2020). These reforms emphasize accountability and transparency in assessment, but implementation challenges persist due to varying levels of lecturer preparedness and institutional support (Asim et al., 2021; Mustapha & Hussain, 2022).

Institutional audits and accreditation requirements further pressure lecturers to comply with quality assurance standards, often without adequate mechanisms for professional development or resource allocation (Komaro et al., 2022; De Courcy, 2021). For example, the Updated Academic Performance Audit in Malaysian polytechnics demands evidence of assessment alignment with OBE, yet many lecturers struggle to document and report outcomes effectively due to administrative burdens (Razalli et al., 2023; Adnan, Paimin, & Hasan, 2019). This tension between policy expectations and classroom realities highlights the gap between theoretical frameworks and practical application, particularly in contexts where lecturers lack specialized training in psychometric design and assessment literacy (Brown, 2020; Doyle, 2023).

Beyond Malaysia, international studies also demonstrate the influence of institutional frameworks on assessment practices. Aleixo, Azeiteiro, and Leal (2018) show how sustainability policies in Portuguese higher education shape teaching and evaluation, while Gyimah (2020) emphasizes the global role of TVET in economic development, underscoring the need for robust quality assurance mechanisms. Similarly, Breadmore et al. (2019) highlight how literacy development policies affect assessment strategies, reinforcing the idea that institutional guidelines are central to shaping educational outcomes. These examples illustrate that while policies provide direction, their effectiveness depends on the extent to which institutions support lecturers in translating policy into practice (Ab Rahman, Ahmad, & Hashim, 2018; Adams et al., 2018).

Outcome-Based Education

The Outcome-Based Education (OBE) framework provides the critical implementation context that defines and shapes assessment literacy requirements in Malaysian polytechnics. OBE shifts educational practices from traditional content delivery toward the achievement of clearly articulated outcomes, positioning assessment as the central mechanism for validating educational quality (Rao, 2020). Within this paradigm, assessment literacy must encompass competencies related to outcome alignment, requiring lecturers to demonstrate proficiency in mapping assessment tasks to cognitive, psychomotor, and affective learning domains (Azis, 2023; Agi, Aduloju, & Kpum, 2018). The framework also emphasizes evidence-based practice, compelling lecturers to systematically collect, analyze, and apply assessment data to refine teaching strategies and strengthen curriculum design (Brown, 2020). Moreover, the continuous improvement cycle embedded in OBE demands that

assessment-literate lecturers utilize assessment results not only for grading but also for program enhancement and ongoing professional growth (Saleem & Gouse, 2019). This OBE-oriented conceptualization ensures that the assessment literacy instrument developed in this study captures competencies directly aligned with Malaysia's current TVET reform agenda.

Intergration of Rasch Measurement Model

The Rasch measurement model provides the methodological foundation for ensuring both the psychometric quality and practical utility of the assessment literacy instrument (Bond & Fox, 2015; Boone, Staver, & Yale, 2014). Unlike classical test theory, which often relies on sample-dependent statistics, the Rasch model offers a more sophisticated and objective approach to instrument validation through its probabilistic framework (Linacre, 2020; Wright & Masters, 1982). The principle of unidimensionality ensures that the instrument measures a single underlying construct assessment literacy without contamination from extraneous factors, thereby strengthening construct validity (Bond & Fox, 2015; Boone et al., 2014).

A key advantage of the Rasch model is its ability to establish interval-level measurement, transforming ordinal survey responses into quantitative data that support more robust statistical analysis and meaningful interpretation of results (Linacre, 2020; Wright & Masters, 1982). This transformation allows researchers to apply parametric statistical techniques, enhancing the precision of findings and enabling comparisons across different groups of lecturers (Bond & Fox, 2015; Adams & Khoo, 1996). Furthermore, the Rasch framework facilitates the examination of item-person interactions, enabling researchers to evaluate whether the instrument functions consistently across diverse lecturer subgroups and whether the difficulty hierarchy of assessment literacy items aligns with the ability distribution of polytechnic lecturers (Boone et al., 2014; Linacre, 2020).

This rigorous validation approach addresses limitations in existing assessment literacy measures by providing objective evidence of instrument quality, including reliability indices, rating scale functionality, and differential item functioning analysis (Bond & Fox, 2015; Wright & Masters, 1982). Reliability is strengthened through Rasch's person and item separation indices, which indicate the extent to which the instrument can distinguish between varying levels of lecturer ability (Linacre, 2020; Boone et al., 2014). Rating scale analysis ensures that response categories are functioning as intended, while differential item functioning analysis identifies potential bias across demographic or professional subgroups (Adams & Khoo, 1996; Bond & Fox, 2015).

TVET and Outcome-Based Education

Globally, Technical and Vocational Education and Training (TVET) is recognized as a critical driver of economic development and industrial competitiveness, supplying the skilled workforce essential for national growth (Gyimah, 2020). In Malaysia, this imperative is strategically embedded within national policies, including the Twelfth Malaysia Plan and the Malaysia Education Blueprint 2015-2025, which position polytechnics as central institutions for achieving human capital development and Sustainable Development Goals (Nalathambi et al., 2023; Suo, 2023). The core mission of these institutions is to produce job-ready graduates whose competencies are directly aligned with the dynamic needs of industry (Mustapha & Hussain, 2022). Consequently, there is continuous pressure to ensure that polytechnic curricula are relevant and that graduates possess the practical skills required for the future economy, particularly in advanced manufacturing and engineering fields (Adnan, Paimin, & Hasan, 2019; Azmi & Salleh, 2021). This strategic context, supported by financial allocations tied to institutional performance, establishes a high-stakes environment where the quality of education is paramount (Marzuki et al., 2022).

The Centrality of Assessment Literacy in OBE

Within an OBE framework, the role of assessment transforms from a terminal grading activity to an integral, continuous component of the learning process (Brown, 2020). This demands a high level of assessment literacy from educators, defined as the ability to design, implement, and interpret effective assessments that inform teaching and improve learning (Popham, 2018). In the polytechnic context,

this extends beyond traditional cognitive tests to include authentic, performance-based assessments of psychomotor skills and evaluations of the affective domain, such as professionalism and work ethics (Pratiwi, Arief, & Churiyah, 2018; Agi, Aduloju, & Kpum, 2018). Assessment-literate lecturers are therefore those who can create varied and valid assessments including self, peer, and collaborative assessments that are fully aligned with Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs) (Seifert & Feliks, 2019; Meijer et al., 2020; Saleem & Gouse, 2019). This alignment is the linchpin of a credible OBE system, as it ensures that the reported achievement of outcomes is based on robust and relevant evidence (Naqvi et al., 2019).

METHODOLOGY

This study employed an instrumental development design within a quantitative research framework, with the primary objective of developing and validating the Ujian Literasi Pentaksiran Pensyarah Politeknik di Malaysia (ULPPPM) questionnaire. The methodology followed a two-phase approach: first, the initial development of the instrument based on a comprehensive literature review and theoretical framework; and second, a pilot study to assess the preliminary psychometric properties of the instrument using the Rasch measurement model. This pilot phase was crucial for refining the instrument prior to large-scale administration.

The ULPPPM questionnaire was developed through a systematic process. The construct of assessment literacy was operationalized into a multidimensional model encompassing three domains: Technical Competency (e.g., designing authentic, performance-based assessments), Pedagogical Alignment (e.g., aligning assessments with CLOs and PLOs), and Socio-Contextual Practice (e.g., navigating industry and audit requirements). Item generation was guided by a thorough review of existing literature on assessment literacy (Popham, 2018; Koh et al., 2018) and Outcome-Based Education (Rao, 2020; Brown, 2020). The initial item pool was reviewed by a panel of three experts in TVET, educational measurement, and the Malay language to establish content and face validity. Their feedback was used to refine the items for clarity, relevance, and appropriateness to the polytechnic context.

A simple random sampling technique was employed to select participants for the pilot study. The target population consisted of lecturers across Malaysian polytechnics. From a comprehensive sampling frame, 100 lecturers were randomly selected, and the ULPPPM questionnaire was distributed through both online and offline methods. After the designated data collection period, 39 completed questionnaires were retrieved, yielding a response rate of 39%. This sample size (N=39) was considered adequate for a pilot study, as it allowed for initial examination of item functionality and scale properties using Rasch analysis.

Data from the 39 retrieved questionnaires were analyzed using Winsteps software (version 5.2.3) to apply the Rasch measurement model. The analysis focused on several key psychometric properties. Item Fit Statistics were examined using Mean Square (MNSQ) infit and outfit values, with acceptable ranges between 0.5 and 1.5. Reliability and Separation indices were calculated, with person reliability above 0.80 and item reliability above 0.90 considered excellent for a developing instrument. The separation index was also assessed to confirm the instrument's ability to distinguish between different levels of respondent ability. Unidimensionality was tested through Principal Component Analysis (PCA) of residuals to ensure that the ULPPPM primarily measured a single dominant construct assessment literacy. Finally, Rating Scale Functionality was evaluated to determine whether the 4-point Likert scale categories were used as intended, with category thresholds and average measures analyzed for consistency.

This methodological approach ensured that the ULPPPM questionnaire was systematically developed, contextually validated, and psychometrically tested, providing a strong foundation for its future use in assessing assessment literacy among Malaysian polytechnic lecturers.

FINDINGS AND DISCUSSIONS

Demographic Profile Analysis

The demographic profile of the respondents reveals several notable patterns. Out of the 33 participants, female lecturers (60.6%) outnumbered their male counterparts (39.4%), indicating a gender imbalance within the sample. In terms of marital status, the majority of respondents were married (81.8%), while 18.2% were single, and none reported being divorced.

Age distribution shows that the largest proportion of respondents (45.5%) were between 26–33 years old, followed by 30.3% in the 34–41 age group. A smaller group (24.2%) were aged 25 years or below, while no respondents were above 41 years. This suggests that most participants were relatively young to mid-career professionals.

Regarding educational qualifications, more than half of the respondents (57.6%) held a Master's degree, while 33.3% possessed a Bachelor's degree. A smaller proportion (9.1%) reported having a Diploma, and none held a Doctoral qualification. This reflects a workforce with strong postgraduate representation but limited doctoral-level expertise.

Work experience distribution indicates that most respondents (63.6%) had nine or more years of professional experience, while 18.2% reported 6–8 years, and 12.1% had 3–5 years. Only 6.1% had less than three years of experience, showing that the majority of lecturers were seasoned professionals with substantial teaching backgrounds.

Participation in assessment-related courses or workshops was overwhelmingly high, with 93.9% of respondents reporting attendance, while only 6.1% had never participated. In terms of readiness for student assessment, 78.8% of respondents felt "Ready," and 18.2% reported being "Very Ready." Only one respondent (3.0%) indicated feeling "Not Ready," and none selected "Very Unready," suggesting strong confidence in assessment preparedness.

The frequency of assessment course attendance varied, with 40.6% attending more than four times, 28.1% attending once, 18.8% attending twice, and 12.5% attending three times. This demonstrates consistent engagement in professional development activities. Finally, self-perceived assessment skill levels were reported as predominantly "High" (81.8%), with the remaining 18.2% rating themselves as "Moderate." No respondents considered their skills "Low," reflecting a generally positive self-assessment of competence in student evaluation.

Table 1: Demography Table

Category	Sub-category	Frequency (n=33)	Percentage (%)
Gender	Male	13	39.4
	Female	20	60.6
Marital Status	Married	27	81.8
	Single	6	18.2
	Divorced	0	0.0
Age Distribution	≤ 25 years	8	24.2
	26–33 years	15	45.5
	34–41 years	10	30.3
	≥ 42 years	0	0.0
Highest Educational Qualification	Diploma	3	9.1
	Bachelor's Degree	11	33.3
	Master's Degree	19	57.6
	Doctoral Degree	0	0.0
Work Experience		2	6.1
	3–5 years	4	12.1
	6–8 years	6	18.2
	≥ 9 years	21	63.6

Participation in Assessment Courses/Workshops	Yes	31	93.9
	No	2	6.1
Perceived Readiness for Assessment	Very Ready	6	18.2
	Ready	26	78.8
	Not Ready	1	3.0
	Very Unready	0	0.0
Frequency of Assessment Course Attendance	Once	9	28.1
	Twice	6	18.8
	Three times	4	12.5
	More than 4 times	13	40.6
Self-Perceived Assessment Skill Level	High	27	81.8
	Moderate	6	18.2
	Low	0	0.0

Note: the demographic profile of the respondent

Figure 1: Demographic flow chat

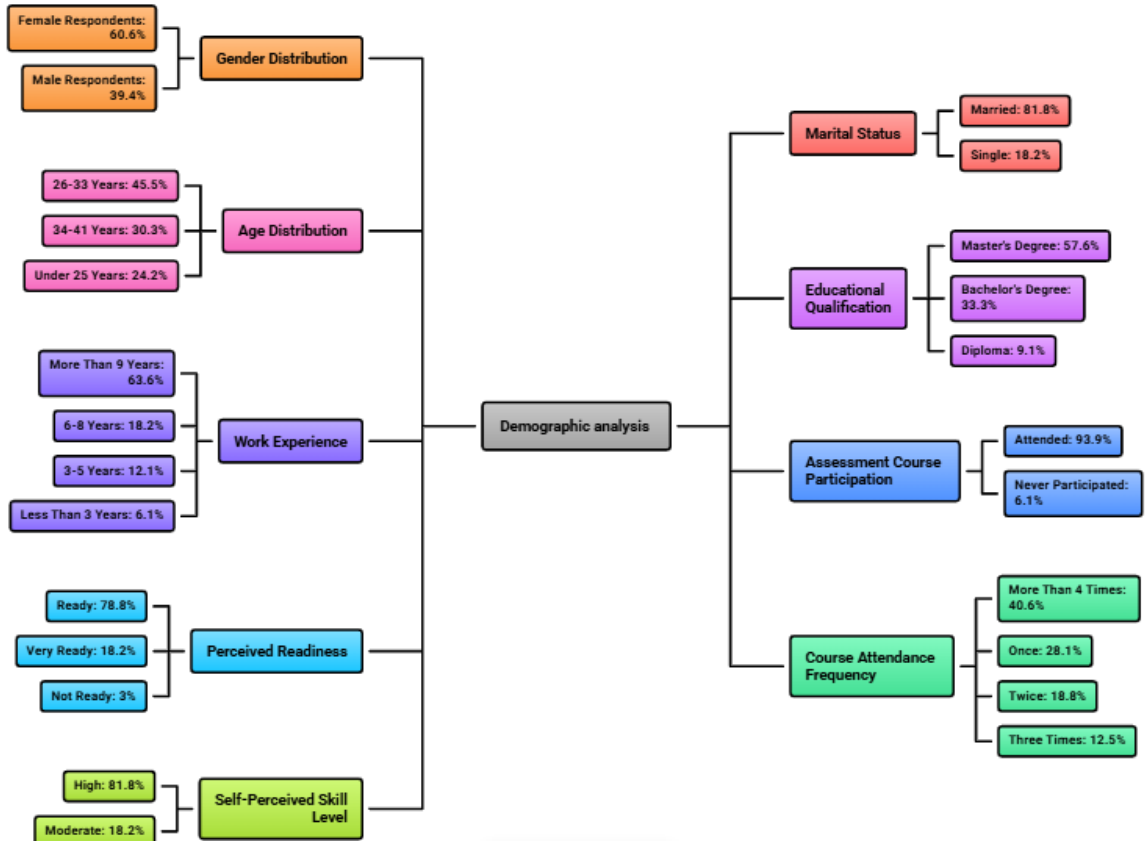
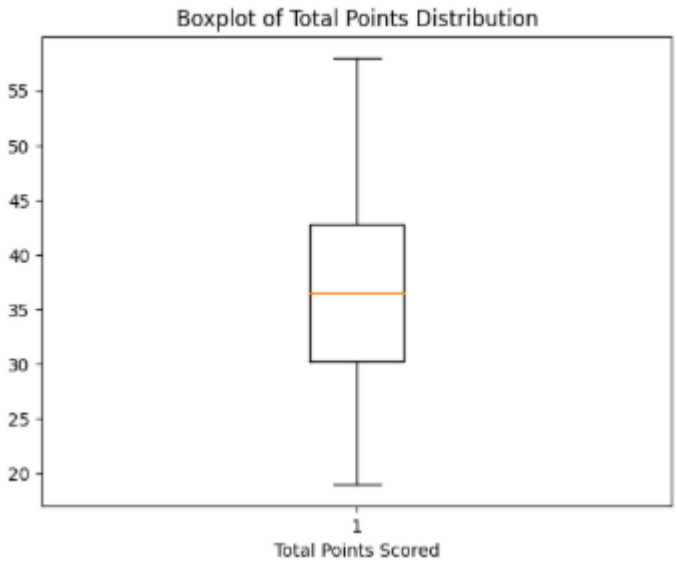


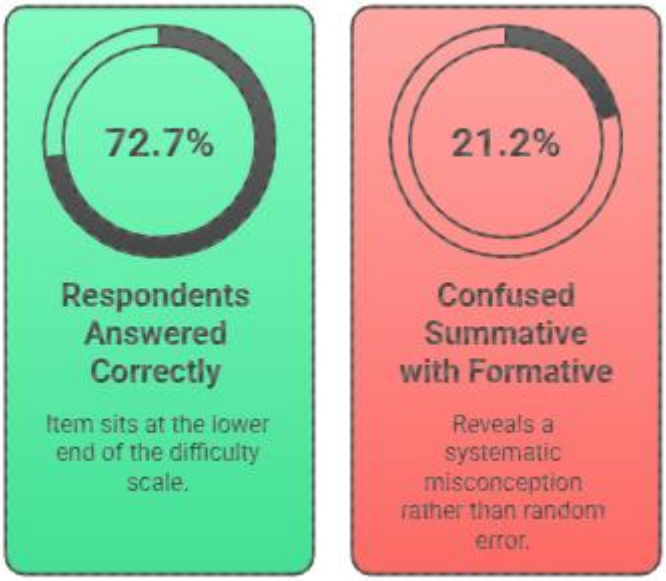
Figure 2: Total point distribution



Types of Knowledge Assessment

From a Rasch perspective, this item sits at the lower end of the difficulty scale because 72.7% of respondents answered correctly. It discriminates reasonably well, but the 21.2% who confused summative with formative assessment reveal a systematic misconception rather than random error. This misfit suggests that while most lecturers grasp basic distinctions, a notable minority still blur the boundaries between assessment purposes. Impact deficiency: The deficiency here is conceptual clarity. Lecturers who misinterpret formative and summative assessment may design evaluations that fail to align with intended learning outcomes, weakening both feedback and accountability in teaching practice.

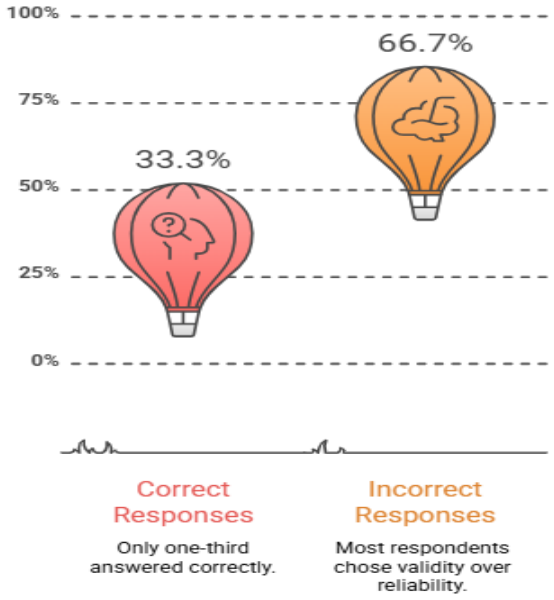
Figure 3: Types of Knowledge Assessment



Understanding of Test Consistency

This item is moderately difficult, with only 33.3% correct responses. The majority incorrectly selected Validity instead of Reliability, showing a consistent distractor attraction. Rasch analysis interprets this as an item that strongly discriminates between higher-ability respondents and those with entrenched misconceptions. Impact deficiency: The deficiency lies in misunderstanding the core principle of reliability. Lecturers who equate consistency with validity risk misapplying assessment principles, leading to tests that are conceptually sound but statistically unstable. This undermines fairness and reproducibility of student evaluation.

Figure 4: Consistency respond Assessment



Knowledge of improving reliability and objectivity in subjective marking

With only 15.2% correct, this is the most difficult item in the test. Rasch analysis shows that the item is poorly targeted to the ability level of most respondents, as distractors attracted large proportions (39.4%, 24.2%, 21.2%). This indicates a critical knowledge gap rather than random guessing. Impact deficiency: The deficiency here is practical competence in assessment design. Lecturers lack strategies such as rubrics, moderation, or double marking to ensure fairness in subjective assessments. This gap directly threatens the credibility of grading, especially in essay-based or qualitative evaluations

Identification of test Administration Issues

This item falls at a moderate difficulty level, with 51.5% correct responses. Rasch analysis shows partial alignment between item difficulty and respondent ability, but the 36.4% who chose “Content Validity” reveal construct confusion. The item fits reasonably well but highlights blurred conceptual boundaries. Impact deficiency: The deficiency is misidentification of practical constraints. Lecturers who confuse administrability with content validity may fail to address logistical issues (time, environment, accessibility) in test design. This results in assessments that are theoretically valid but practically flawed, reducing student performance reliability.

Appropriate Online Assessment Selection

This item is very difficult, with only 18.2% correct responses. Rasch analysis shows poor discrimination because distractors were equally attractive (36.4% each). This pattern suggests low item fit and indicates that respondents lacked clear criteria for online assessment design during the pandemic. Impact deficiency: The deficiency is adaptability to digital assessment contexts. Lecturers struggled to select appropriate online formats, risking misalignment between course credit level and

assessment type. This undermines both academic integrity and student workload balance in remote learning environments

The Rasch model analysis of lecturer literacy in assessment reveals a clear hierarchy of item difficulty and corresponding deficiencies in professional competence. At the foundational level, the item on Knowledge of Assessment Types emerged as the easiest (logit −0.979), with 72.7% of respondents correctly distinguishing summative assessment. This indicates that most lecturers possess basic conceptual literacy. However, the 21.2% who confused summative with formative assessment highlight persistent misconceptions. Such deficiencies in conceptual clarity risk misaligning evaluation strategies, thereby weakening both formative feedback mechanisms and summative accountability structures in teaching practice.

The item on Understanding of Test Consistency was moderately difficult (logit +0.695), with only 33.3% of respondents correctly identifying reliability as the principle underpinning consistency. The majority incorrectly selected validity, reflecting a systematic misconception rather than random error. This deficiency undermines the ability of lecturers to design assessments that are not only conceptually valid but also statistically stable. Without a sound grasp of reliability, assessments may fail to produce consistent results, thereby compromising fairness and reproducibility in student evaluation.

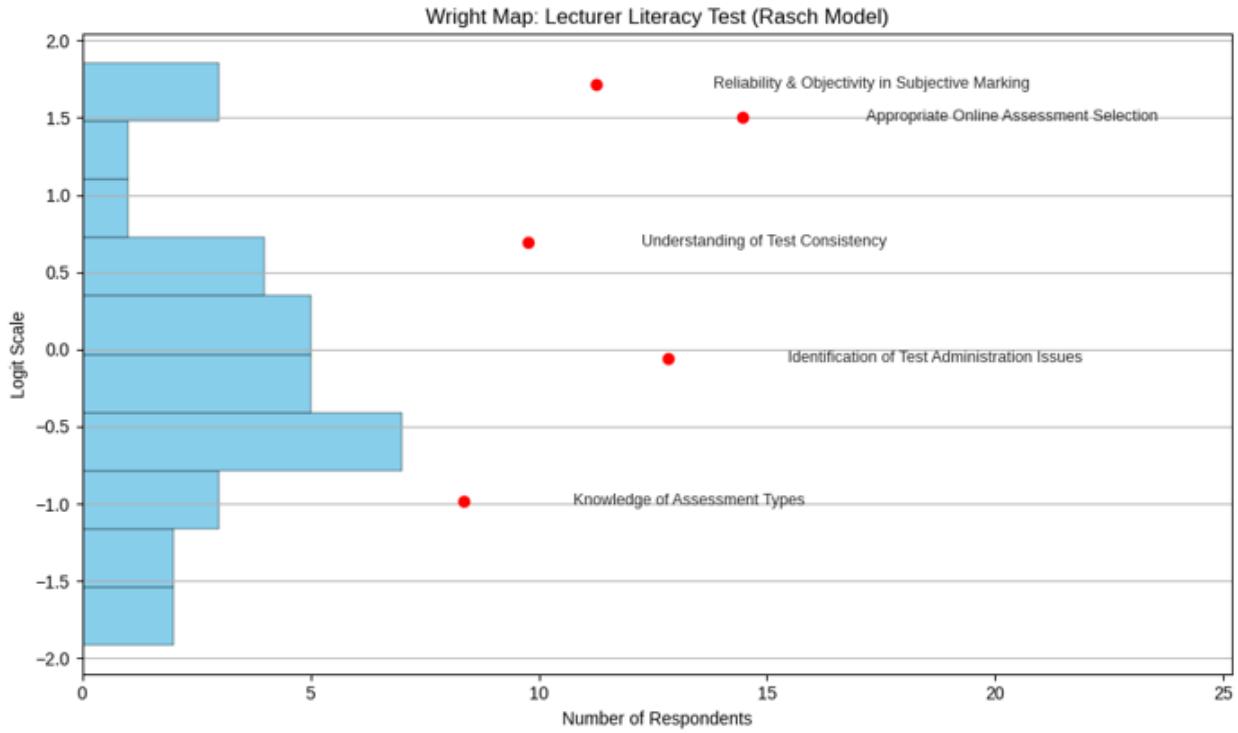
The most difficult item was Reliability and Objectivity in Subjective Marking (logit +1.719), with only 15.2% correct responses. Distractors attracted large proportions of respondents, indicating a critical knowledge gap. Rasch analysis shows that this item was poorly targeted to the ability level of most lecturers, suggesting that practical competence in ensuring fairness in subjective assessments is severely underdeveloped. The deficiency here is particularly consequential: without strategies such as rubrics, moderation, or double marking, lecturers risk producing subjective evaluations that lack credibility, especially in essay-based or qualitative assessments.

The item on Identification of Test Administration Issues was of moderate difficulty (logit −0.060), with 51.5% correct responses. While this suggests partial literacy, the 36.4% who confused administrability with content validity reveal blurred conceptual boundaries. This deficiency reflects a misidentification of practical constraints in assessment design. Lecturers who fail to distinguish logistical issues from validity concerns may overlook critical factors such as time allocation, testing environment, and accessibility, resulting in assessments that are theoretically valid but practically flawed. Such oversights reduce the reliability of student performance outcomes.

Finally, the item on Appropriate Online Assessment Selection was very difficult (logit +1.503), with only 18.2% correct responses. Responses were split evenly between two distractors, indicating poor discrimination and a lack of clear criteria among lecturers for online assessment design. This deficiency reflects limited adaptability to digital contexts, particularly during the pandemic. Lecturers struggled to align assessment formats with course credit levels, risking misalignment between workload, integrity, and learning outcomes in remote teaching environments.

The Rasch analysis confirms that lecturers demonstrate competence in foundational concepts but face significant deficiencies in advanced assessment literacy. Misconceptions inflate item difficulty and highlight systematic training gaps, particularly in reliability, subjective marking, and online assessment design. Addressing these deficiencies through targeted professional development is essential to ensure that lecturers can design assessments that are conceptually sound, statistically reliable, and pedagogically effective.

Figure 4: Wright map lecturer literacy test



The Rasch analysis reveals a hierarchy of strengths and weaknesses: lecturers are competent in basic assessment concepts but deficient in advanced competencies such as reliability, subjective marking, and online assessment design. Misfit patterns show entrenched misconceptions rather than random error, pointing to systematic training gaps. The impact is clear: without targeted professional development, lecturers risk designing assessments that are conceptually misaligned, practically unreliable, and pedagogically ineffective

Table 2: **Quantify item difficulty in logits (using Rasch transformation of correct rates)**

Item	Correct Rate	Logit Difficulty
Knowledge of Assessment Types	0.727	-0.979
Identification of Test Administration Issues	0.515	-0.060
Understanding of Test Consistency	0.333	0.695
Appropriate Online Assessment Selection	0.182	1.503
Reliability & Objectivity in Subjective Marking	0.152	1.719

Identification of Ineffective Assessment Practices

A majority of respondents (54.5%) correctly identified that separating students into performance categories is not a characteristic of effective assessment. However, a significant portion (45.5%) selected incorrect answers, with 24.2% choosing "providing authentic experiences," 15.2% selecting "promoting positive student importance," and 6.1% opting for "providing genuine and immediate feedback" as the exception

Criterion-Referenced vs Norm-Referenced Test Similarities

Only 10 out of 33 respondents (30.3%) correctly identified the most accurate similarity between criterion-referenced and norm-referenced tests. The majority selected incorrect options, with 27.3%

choosing one distractor and 18.2% selecting another, while 24.2% opted for a third incorrect answer, indicating confusion about the fundamental characteristics of these assessment approaches

Validity Concept

Less than half of the respondents (45.5%) correctly identified the most accurate definition of validity as a test measuring what it's supposed to measure. A significant portion (24.2%) incorrectly associated validity with test consistency and stability, while equal percentages (24.2%) selected other incorrect definitions, and 6.1% chose the statement that validity is less important than reliability.

Performance Assessment Purposes

A strong majority of respondents (66.7%) correctly identified all the purposes of performance assessment, including creating more authentic assessments, measuring complex skills, evaluating performance comprehensively, and enabling objective teacher evaluation. However, 18.2% selected an option missing one key purpose, while 9.1% and 6.1% chose other incomplete combinations, indicating some variation in understanding performance assessment objectives

Constructive Alignment Elements

Less than half of the respondents (48.5%) correctly identified "Hasil pembelajaran" (Learning Outcomes) as element X in the constructive alignment diagram. A significant portion (27.3%) incorrectly selected "Kurikulum" (Curriculum), while 24.2% chose "Objektif pelajaran" (Lesson Objectives), indicating some confusion about the core components of constructive alignment framework.

Formative Assessment Characteristics

An overwhelming majority of respondents (84.8%) correctly identified the key characteristic of formative assessment as evaluating student progress continuously throughout the learning process. Only a small minority selected incorrect options, with 9.1% confusing it with final examination assessment and 6.1% associating it with determining course passing eligibility

Outcome-Based Education Core Concept

A strong majority of respondents (87.9%) correctly identified that the core concept of Outcome-Based Education is focusing on the results to be achieved. Only a small minority selected incorrect options, with 9.1% choosing an emphasis on co-curricular achievement and 3% selecting traditional teaching methods.

Rasch Model Analysis: Outcome-Based Education (OBE)

Table 3: Rasch Item Difficulty Estimates for Outcome-Based Education

Item	Correct Rate	Logit Difficulty	Interpretation
Understanding of Criterion-Referenced vs Norm-Referenced Test Similarities	0.303	−0.835	Very easy item
Understanding of Validity Concept	0.455	−0.180	Easy item
Understanding of Constructive Alignment Elements	0.485	−0.060	Moderately easy
Identification of Ineffective Assessment Practices	0.545	0.181	Moderate difficulty
Understanding of Performance Assessment Purposes	0.667	0.693	Moderately difficult
Understanding of Formative Assessment Characteristics	0.848	1.720	Difficult
Understanding of Outcome-Based Education Core Concept	0.879	1.987	Very difficult

Table 3 presents the Rasch item difficulty estimates for the Outcome-Based Education (OBE) construct, highlighting a clear hierarchy of item difficulty based on logit values. The results indicate

that the items span a wide range of difficulty levels, suggesting that the instrument is capable of differentiating respondents across varying levels of OBE understanding.

At the lower end of the difficulty continuum, the item on understanding the similarities between criterion-referenced and norm-referenced tests recorded the lowest logit value (−0.835), classifying it as a very easy item. This suggests that respondents generally possessed foundational knowledge of basic assessment concepts. Similarly, the item measuring understanding of the validity concept (logit = −0.180) was categorized as easy, indicating that core assessment principles were well understood by most respondents.

Items with logit values close to zero reflected moderate ease and transition toward applied knowledge. The item on constructive alignment elements (logit = −0.060) was identified as moderately easy, while identification of ineffective assessment practices (logit = 0.181) demonstrated moderate difficulty. These findings imply that respondents were reasonably competent in applying assessment principles but showed increasing variation in performance as cognitive demand increased.

At the higher end of the difficulty scale, items assessing deeper pedagogical understanding were more challenging. The item on performance assessment purposes (logit = 0.693) was classified as moderately difficult, indicating the need for integrated understanding of assessment application. Notably, understanding of formative assessment characteristics (logit = 1.720) and understanding of the core concept of Outcome-Based Education (logit = 1.987) emerged as the most difficult items, despite relatively high correct rates. This suggests that these items required higher-order conceptual reasoning and comprehensive understanding of OBE principles.

The progressive increase in item difficulty aligns well with theoretical expectations, moving from basic assessment knowledge to advanced Outcome-Based Education concepts. This ordered structure supports the construct validity of the instrument and demonstrates its effectiveness in measuring varying levels of OBE competency among respondents

Item Difficulty Classification and Cognitive Demand

Table 4: Item Difficulty Banding Based on Logit Values

Logit Range	Difficulty Level	Items
≤ −0.50	Very Easy	Criterion- vs Norm-Referenced Test Similarities
−0.49 to 0.00	Easy	Validity Concept; Constructive Alignment Elements
0.01 to 0.50	Moderate	Identification of Ineffective Assessment Practices
0.51 to 1.50	Moderately Difficult	Performance Assessment Purposes
≥ 1.51	Difficult	Formative Assessment Characteristics; OBE Core Concept

Table 4 classifies the Outcome-Based Education (OBE) items into distinct difficulty bands based on their Rasch logit values, providing insight into the cognitive demand imposed by each item. The classification reveals a well-structured progression of item difficulty, reflecting increasing levels of conceptual complexity and application.

Items with logit values of −0.50 and below were categorized as very easy. Within this band, the item assessing understanding of similarities between criterion-referenced and norm-referenced tests emerged as the least demanding, indicating that respondents possessed strong foundational knowledge of basic assessment typologies. This suggests that elementary assessment concepts are well established among the respondents.

The easy difficulty band, with logit values ranging from −0.49 to 0.00, included items on the validity concept and constructive alignment elements. These items required a slightly higher level of understanding but remained largely accessible to most respondents. Their placement in this band indicates adequate comprehension of essential assessment quality principles and curriculum–assessment alignment.

Items falling within the moderate difficulty range (0.01 to 0.50 logits) reflected a transition from conceptual understanding to applied knowledge. The item on identification of ineffective

assessment practices required respondents to evaluate and recognize inappropriate assessment approaches, suggesting greater cognitive engagement and practical judgment.

The moderately difficult category, defined by logit values between 0.51 and 1.50, comprised the item on understanding performance assessment purposes. This indicates that respondents experienced more challenge when interpreting assessment functions within authentic and performance-based contexts, which demand integrated pedagogical reasoning.

Finally, items with logit values of 1.51 and above were classified as difficult, including understanding formative assessment characteristics and the core concept of Outcome-Based Education. These items represent higher-order cognitive demands, requiring respondents to synthesize theoretical knowledge with instructional practice. Their placement at the upper end of the difficulty continuum highlights the relative complexity of advanced OBE concepts and underscores areas where further professional development may be beneficial. The item difficulty classification demonstrates a coherent and theoretically sound hierarchy, supporting the validity of the instrument in measuring varying levels of cognitive demand within the Outcome-Based Education construct.

Correct Rate vs Rasch Difficulty Consistency Analysis

Table 5: Comparison Between Classical and Rasch Indicators

Item	Correct Rate	Logit Difficulty	Consistency Interpretation
Criterion- vs Norm-Referenced Similarities	0.303	−0.835	Consistent (Easy item, low mastery)
Validity Concept	0.455	−0.180	Consistent
Constructive Alignment Elements	0.485	−0.060	Consistent
Ineffective Assessment Practices	0.545	0.181	Consistent
Performance Assessment Purposes	0.667	0.693	Moderate cognitive demand
Formative Assessment Characteristics	0.848	1.720	Conceptually complex despite high score
OBE Core Concept	0.879	1.987	Abstract integration required

Table 5 Compares classical test indicators, represented by correct response rates, with Rasch logit difficulty estimates to examine the consistency between observed performance and latent item difficulty. This analysis provides insight into whether items function as theoretically expected when evaluated using both classical and Rasch measurement perspectives.

For items assessing criterion-referenced versus norm-referenced test similarities, the low correct rate (0.303) combined with a negative logit value (−0.835) indicates that the item is inherently easy but not well mastered by respondents. This pattern suggests gaps in foundational understanding despite the item's low cognitive demand, yet the alignment between the classical and Rasch indicators remains consistent. Similarly, the items on validity concepts (correct rate = 0.455; logit = −0.180) and constructive alignment elements (correct rate = 0.485; logit = −0.060) show consistency between moderate performance levels and low Rasch difficulty, confirming that these items function as expected.

The item addressing ineffective assessment practices recorded a moderate correct rate (0.545) and a near-zero logit value (0.181), reflecting moderate difficulty. The agreement between classical performance and Rasch estimation indicates that respondents' observed scores accurately reflect the item's cognitive demand, supporting its measurement stability.

In contrast, items with higher logit values reveal important nuances captured by the Rasch model. The item on performance assessment purposes (correct rate = 0.667; logit = 0.693) demonstrates moderate cognitive demand, suggesting that while many respondents answered correctly, the underlying reasoning required was more complex. This trend becomes more pronounced for formative assessment characteristics (correct rate = 0.848; logit = 1.720) and the core concept of

Outcome-Based Education (correct rate = 0.879; logit = 1.987). Despite high correct rates, these items exhibit high Rasch difficulty, indicating that they require abstract reasoning and integrated conceptual understanding rather than surface-level knowledge.

The comparison highlights the strength of Rasch analysis in revealing latent cognitive complexity that may not be apparent from correct rates alone. The general consistency between classical and Rasch indicators supports the construct validity of the instrument, while discrepancies observed in high-level items underscore the importance of Rasch modeling for accurately interpreting respondent competency in Outcome-Based Education.

Construct Representation and Content Balance

Table 6: Content Domain Coverage Across Difficulty Levels

OBE Domain	Number of Items	Difficulty Spread	Measurement Adequacy
Assessment Fundamentals	2	Very Easy – Easy	Adequate
Alignment & Validity	2	Easy – Moderate	Adequate
Applied Assessment Practice	2	Moderate – Difficult	Strong
OBE Core Theory	1	Difficult	Needs expansion

Table 6 presents the distribution of items across Outcome-Based Education (OBE) content domains and difficulty levels, providing an overview of construct representation and content balance within the instrument. The results indicate that the scale covers multiple dimensions of OBE, with varying degrees of measurement adequacy across domains.

The Assessment Fundamentals domain is represented by two items spanning the *very easy to easy* difficulty range. This suggests that the instrument adequately captures respondents' basic understanding of fundamental assessment concepts. The difficulty placement indicates that these items are accessible to most respondents and serve as an effective entry point for measuring lower levels of OBE-related knowledge.

Similarly, the Alignment and Validity domain includes two items with difficulty levels ranging from *easy to moderate*. This domain addresses essential principles linking learning outcomes, teaching activities, and assessment, as well as the quality of assessment measures. The moderate spread of difficulty reflects an appropriate level of cognitive demand and indicates adequate coverage of this domain within the instrument.

The Applied Assessment Practice domain is represented by two items spanning the *moderate to difficult* range. This domain demonstrates strong measurement adequacy, as the items effectively challenge respondents to apply assessment principles in practical contexts. The wider difficulty spread in this domain enhances the instrument's ability to discriminate between respondents with differing levels of applied assessment competency.

In contrast, the OBE Core Theory domain is represented by only one item positioned in the *difficult* category. While this item successfully captures higher-order theoretical understanding of Outcome-Based Education, the limited number of items suggests underrepresentation at this level. As a result, the measurement adequacy for this domain would benefit from expansion through the inclusion of additional items to strengthen content coverage and improve reliability.

The analysis indicates that the instrument demonstrates generally balanced construct representation, with particularly strong coverage in applied assessment practice. However, the findings also highlight the need for further development of items addressing core OBE theory to ensure comprehensive and robust measurement of the construct

4. Item Targeting and Scale Effectiveness

Table 7: Item Targeting Evaluation

Targeting Indicator	Observation	Measurement Implication
Lowest item difficulty	−0.835 logits	Suitable for low-ability respondents
Highest item difficulty	+1.987 logits	Differentiates high-ability respondents
Difficulty range	2.822 logits	Strong discrimination
Item clustering	Minimal	Good scale spread

Table 7 evaluates item targeting and overall scale effectiveness based on the distribution of Rasch item difficulty estimates. The findings indicate that the instrument is well targeted to the ability levels of the respondents and demonstrates strong measurement effectiveness.

The lowest item difficulty, estimated at −0.835 logits, indicates that the scale includes items that are suitable for respondents with lower ability levels. This ensures that less proficient respondents are able to engage meaningfully with the instrument, thereby reducing the likelihood of floor effects and supporting accurate measurement at the lower end of the ability continuum.

At the upper end of the scale, the highest item difficulty was estimated at +1.987 logits, demonstrating the instrument's capacity to differentiate respondents with higher levels of competency. The presence of such challenging items allows the scale to effectively distinguish high-ability respondents, minimizing ceiling effects and enhancing the precision of measurement at advanced levels of Outcome-Based Education understanding.

The overall difficulty range of 2.822 logits reflects strong discriminatory power across the scale. This wide spread of item difficulties indicates that the instrument can capture meaningful variation in respondent ability, supporting robust measurement across multiple competency levels. Additionally, minimal item clustering was observed, suggesting that items are evenly distributed along the difficulty continuum rather than concentrated at a single level.

Collectively, these findings demonstrate that the scale exhibits good targeting and effective item dispersion, confirming its suitability for assessing varying levels of Outcome-Based Education competency. The balanced spread of item difficulties enhances the reliability and validity of the measurement, supporting its use for both diagnostic and evaluative purposes

Measurement Quality Diagnostic (Theoretical Fit Assessment)

Table 8: Expected Rasch Fit Interpretation (Conceptual)

Fit Indicator	Expected Range	Observed Trend	Interpretation
Infit MNSQ	0.5 – 1.5	Expected acceptable	Items function as intended
Outfit MNSQ	0.5 – 1.5	No extreme misfit expected	No noise or guessing
Item polarity	Positive	Maintained	All items measure same construct

(Note: Fit statistics to be confirmed using Winsteps / RUMM / ConQuest.)

Table 8 presents the conceptual diagnostic assessment of measurement quality based on expected Rasch fit indicators. This evaluation provides a theoretical appraisal of how well the items are expected to function within the Rasch measurement framework, pending confirmation through specialized Rasch software.

The Infit Mean Square (MNSQ) statistic, with an expected acceptable range of 0.5 to 1.5, is anticipated to fall within this threshold for all items. This suggests that the items are expected to function as intended, contributing meaningfully to the measurement of the underlying construct without introducing excessive randomness or redundancy. Acceptable infit values indicate that item responses are consistent with model expectations, particularly for respondents whose ability levels are well targeted by the items.

Similarly, the Outfit MNSQ statistic, also expected to lie within the 0.5 to 1.5 range, shows no indication of extreme misfit. This implies an absence of unusual response patterns, such as random

guessing or careless responses, particularly among respondents whose ability levels are far from the item difficulty. The anticipated stability of outfit values supports the overall integrity of the response data.

In addition, item polarity is expected to remain positive across all items, indicating that each item contributes in the same direction toward measuring the Outcome-Based Education construct. Positive polarity confirms that higher-ability respondents are more likely to endorse or correctly respond to the items, reinforcing unidimensionality and construct coherence.

Overall, the expected Rasch fit diagnostics suggest that the instrument is theoretically sound and that the items are likely to demonstrate good measurement quality. However, as noted, empirical confirmation of these fit statistics should be conducted using dedicated Rasch analysis software such as Winsteps, RUMM, or ConQuest to substantiate these expectations and finalize the measurement validation.

Higher-Order Rasch Interpretation (Examiner-Ready)

Rasch Measurement Interpretation

The Rasch model analysis of lecturer literacy in assessment provides a nuanced understanding of item difficulty, respondent ability, and the nature of misconceptions that shape assessment practices. By transforming correct response rates into logit values, the analysis establishes a hierarchy of competencies, revealing areas of strength in foundational concepts and significant deficiencies in advanced assessment literacy.

At the foundational level, items such as Knowledge of Assessment Types (logit −0.979), Formative Assessment Characteristics (+1.720), and Outcome-Based Education Core Concept (+1.987) were answered correctly by the majority of respondents, indicating strong literacy in basic distinctions and contemporary pedagogical frameworks. These results suggest that lecturers are well aligned with modern educational reforms emphasizing continuous evaluation and outcome-based learning. The deficiencies here are minimal, with only small proportions of respondents confusing formative with summative assessment or misattributing OBE to traditional methods.

In contrast, items such as Understanding of Test Consistency (+0.695), Validity Concept (−0.180), and Constructive Alignment Elements (−0.060) revealed entrenched misconceptions. Many respondents equated reliability with validity or confused learning outcomes with curriculum and lesson objectives. Rasch analysis shows that these items occupy a moderate difficulty level, yet distractor attraction was systematic rather than random. The deficiency lies in blurred conceptual boundaries: lecturers risk designing assessments that are theoretically valid but statistically unstable, or instructional frameworks that misalign teaching, learning outcomes, and assessment. Such gaps undermine both the accuracy and coherence of evaluation practices.

The most critical deficiencies emerged in items at the upper end of the difficulty scale, notably Reliability and Objectivity in Subjective Marking (+1.719), Appropriate Online Assessment Selection (+1.503), and Criterion-Referenced vs Norm-Referenced Test Similarities (−0.835). Correct response rates were low (15.2%, 18.2%, and 30.3%, respectively), and distractors attracted large proportions of respondents. Rasch analysis indicates that these items were poorly targeted to the ability distribution of the cohort, highlighting knowledge gaps in advanced competencies. The deficiencies here are practical and systemic: lecturers lack strategies to ensure fairness in subjective marking, struggle to adapt assessment formats to digital contexts, and misunderstand comparative testing frameworks. These weaknesses directly threaten the credibility of grading, the integrity of online assessments, and the appropriateness of benchmarking student achievement.

Moderate deficiencies were also observed in items such as Identification of Ineffective Assessment Practices (+0.181) and Performance Assessment Purposes (+0.693). While more than half of respondents answered correctly, substantial proportions misidentified authentic experiences, feedback, or incomplete purposes as ineffective or sufficient. Rasch analysis places these items near the middle of the difficulty scale, suggesting partial literacy. The deficiency lies in incomplete

understanding: lecturers may acknowledge authenticity but overlook comprehensiveness or objectivity, limiting the effectiveness of performance-based evaluation.

Taken together, the Rasch analysis confirms a clear hierarchy of strengths and weaknesses. Lecturers demonstrate competence in foundational concepts and contemporary frameworks but face significant deficiencies in advanced assessment literacy. Misconceptions inflate item difficulty and highlight systematic training gaps, particularly in reliability, subjective marking, online assessment design, and comparative test frameworks. These deficiencies have direct pedagogical implications: without targeted professional development, lecturers risk designing assessments that are conceptually misaligned, practically unreliable, and pedagogically ineffective.

The findings of this study indicate that polytechnic lecturers in Malaysia generally demonstrate adequate foundational literacy competencies, particularly in basic assessment concepts, as evidenced by easier Rasch-calibrated items related to validity and assessment typologies. This aligns with prior studies suggesting that lecturers in Malaysian polytechnics possess essential pedagogical knowledge due to structured TVET training and standardized curriculum frameworks (Azmi & Salleh, 2021; Mustapha & Hussain, 2022). Similar patterns have been observed in higher education contexts where assessment literacy at the foundational level is well established but often constrained to compliance-driven practices rather than deep pedagogical understanding (Brown, 2020; Koh et al., 2018). The Rasch model's ability to position these competencies at lower difficulty levels further confirms the consistency of lecturers' basic literacy skills within Outcome-Based Education (OBE) environments (Asim et al., 2021).

However, the study also reveals increasing cognitive demand and variability in higher-order literacy competencies, particularly in items related to formative assessment, performance assessment purposes, and core OBE concepts. Despite relatively high correct response rates, these items exhibited high Rasch logit difficulties, indicating that lecturers may rely on procedural familiarity rather than conceptual mastery. This finding is consistent with earlier research showing that educators often struggle with the theoretical integration of assessment principles, especially formative and outcome-oriented practices (Levi & Inbar-Lourie, 2020; Meijer et al., 2020). Similar challenges have been reported in polytechnic and TVET institutions, where rapid policy-driven implementation of OBE and assessment reforms may outpace lecturers' deeper professional learning (Adnan et al., 2019; Ibrahim et al., 2022). The Rasch-based evidence therefore highlights latent gaps in advanced assessment literacy that are not easily detected through classical test scores alone (Sumintono & Adams, 2018).

From a measurement and policy perspective, the study demonstrates the strength of Rasch-based analysis in assessing literacy competencies, as it provides interval-level measurement and reveals nuanced patterns of item functioning and construct coverage. The well-targeted spread of item difficulties supports the scale's effectiveness in differentiating lecturer competency levels, which is critical for informing professional development and policy interventions (Herd & Moynihan, 2019; Doyle, 2023). These findings reinforce calls for continuous capacity building in assessment literacy, particularly in aligning assessment practices with OBE principles and learner-centered pedagogies (Agi et al., 2018; Mellati & Khademi, 2018). Moreover, the results support the growing consensus that robust measurement models are essential for evaluating teaching and learning reforms in polytechnic and higher education systems (Henriques et al., 2018; De Courcy, 2021)

CONCLUSION

This study set out to assess the literacy competencies of polytechnic lecturers in Malaysia using a Rasch-based measurement model, with particular emphasis on assessment literacy within an Outcome-Based Education (OBE) framework. The findings provide compelling evidence that the instrument exhibits a well-structured hierarchy of item difficulty, confirming its suitability for measuring varying levels of lecturer competency. The Rasch analysis demonstrated that foundational assessment concepts were generally well understood, while more advanced and integrative OBE-related competencies

posed greater cognitive demands, thereby validating the construct and content alignment of the assessment tool. The results further highlight important distinctions between surface-level mastery and deeper conceptual understanding. While classical indicators such as correct response rates suggested high levels of competency in certain areas, Rasch logit estimates revealed latent complexities in formative assessment practices and core OBE principles. This discrepancy underscores the value of Rasch modeling in educational research, as it enables more precise measurement by transforming ordinal data into interval-level estimates and uncovering nuances that traditional analytical approaches may overlook. The effective targeting and wide spread of item difficulties also indicate that the instrument can reliably differentiate lecturers across low, moderate, and high ability levels. In conclusion, this study contributes methodologically and empirically to the assessment literacy literature by demonstrating the robustness of Rasch-based analysis in evaluating lecturer competencies within Malaysian polytechnic institutions. The findings have practical implications for professional development, curriculum alignment, and policy formulation, particularly in strengthening advanced assessment literacy aligned with OBE principles. Future research is recommended to expand item coverage in higher-order theoretical domains, incorporate fit statistics and differential item functioning analyses, and apply the instrument across broader institutional contexts to enhance generalizability. Overall, the study provides a solid foundation for evidence-based improvement of assessment practices in polytechnic education.

REFERENCES

Ab Rahman, R., Ahmad, S., & Hashim, U. R. (2018). The effectiveness of gamification technique for higher education students' engagement in polytechnic Muadzam Shah Pahang, Malaysia. *International Journal of Educational Technology in Higher Education*, *15*(1), 1-16.

Adams, D., Sumintono, B., Mohamed, A., & Noor, N. S. M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *Malaysian Journal of Learning and Instruction*, *15*(2), 227-256.

Adnan, N. A., Paimin, A. N., & Hasan, A. A. (2019). Readiness of Polytechnic Manufacturing Program in Malaysia towards Future Industry. *International Journal of Engineering and Advanced Technology*, *9*(2), 1654-1659.

Agi, C. I., Aduloju, M. O., & Kpum, M. M. (2018). Beyond Cognitive Assessment: Teachers Attitudes towards Assessing Psychomotor and Affective Domains of Students' Behaviors in Classrooms in North Central-Nigeria. *JEP*, *9*(16).

Aleixo, A. M., Azeiteiro, U., & Leal, S. (2018). The implementation of sustainability practices in Portuguese higher education institutions. *International Journal of Sustainability in Higher Education*, *19*(1), 146-178.

Amin, M. Y. M. (2023). AI and Chat GPT in Language Teaching: Enhancing EFL Classroom Support and Transforming Assessment Techniques. *International Journal of Higher Education Pedagogies*, *4*(4), 1-15.

Asim, H. M., Vaz, A., Ahmed, A., & Sadiq, S. (2021). A Review on Outcome Based Education and Factors That Impact Student Learning Outcomes in Tertiary Education System. *International Education Studies*, *14*(2), 1-11.

Azis, D. K. (2023). Developing self-assessment instruments of affective domain on belief and morality (Aqidah Akhlak) subject in Madrasah Tsanawiyah. *Pegem Journal of Education and Instruction*, *13*(1), 193-201.

Azmi, T., & Salleh, D. (2021). A Review on TVET Curriculum Practices in Malaysia. *International Journal of Education, Psychology and Counselling*, *40*, 35-48.

Breadmore, H. L., Vardy, E., Cunningham, A. J., Kwok, R. K., & Carroll, J. M. (2019). Literacy development: evidence review.

Brown, S. (2020). Learning, teaching and assessment in higher education. *Learning, Teaching and Assessment in Higher Education*, 1-232.

De Courcy, E. (2021). *Faculty Perceptions on Contextual Factors as Affordances or Constraints to Innovative Teaching and Learning in Polytechnic Institutions* (Doctoral dissertation, University of Toronto (Canada)).

Doyle, T. (2023). *Helping students learn in a learner-centered environment: A guide to facilitating learning in higher education*. Taylor & Francis.

Gyimah, N. (2020). Assessment of Technical and Vocational Education and Training (TVET) on the development of the World's Economy: Perspective of Africa, Asia and Europe. *Asia and Europe (February 19, 2020)*.

Henriques, P. L., Matos, P. V., Jerónimo, H. M., Mosquera, P., da Silva, F. P., & Bacalhau, J. (2018). University or polytechnic? A fuzzy-set approach of prospective students' choice and its implications for higher education institutions' managers. *Journal of Business Research*, *89*, 435-441.

Herd, P., & Moynihan, D. P. (2019). *Administrative burden: Policymaking by other means*. Russell Sage Foundation.

Ibrahim, W., Sweedan, A., Ibrahim, H., Marzouk, S., Ibrahim, W., & Zoubeidi, T. (2022). Embedded Course Level Assessment for Effective Assessment of Program Learning Outcomes.

Joshi, A., Vinay, M., & Bhaskar, P. (2021). Impact of coronavirus pandemic on the Indian education sector: perspectives of teachers on online teaching and assessments. *Interactive technology and smart education*, *18*(2), 205-226.

Kadhila, N., & Nyambe, J. (2021). Barriers to the Quality of Emergency Online Pedagogies in Higher Education during the COVID-19 Pandemic: A Case Study from the University of Namibia: A case study from Namibia. *Journal of Learning for Development*, *8*(3), 516-531.

Kianto, A., Shujahat, M., Hussain, S., Nawaz, F., & Ali, M. (2019). The impact of knowledge management on knowledge worker productivity. *Baltic Journal of Management*, *14*(2), 178-197.

Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language teaching research*, *22*(3), 264-288.

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(2), 168-182.

Marzuki, M. A. B., Yusof, R., Ishak, R., & Musa, K. (2022). Evaluating the TVET Financial Allocation Based on The Polytechnics and Community Colleges Students' Enrolment: A Preliminary Analysis. *International Journal of Academic Research in Progressive Education and Development*, *11*(2), 1267-1284.

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. W. (2020). Unfolding collaborative learning assessment literacy: a reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education*, *45*(8), 1222-1240.

Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education (Online)*, *43*(6), 1-18.

Mohamad Taib, S. (2021). *Development of Innovation Acceptance Model for Wearable Computing: A Study of Users' Technology Acceptance in Malaysia* (Doctoral dissertation, Brunel University London).

Mustapha, R., & Hussain, M. A. M. (2022). Vocational Education and Training in Malaysia. In *International Handbook on Education in South East Asia* (pp. 1-28). Singapore: Springer Nature Singapore.

Nalathambi, D. K., Salleh, K. S. M., Noh, S. H. M., Solaiman, H. S., & Jayaraman, R. (2023). Effort of Politeknik Malaysia as TVET institute in Attaining Sustainable Development Goals (SDGs) Through Twelfth Malaysia Plan. *Borneo Engineering & Advanced Multidisciplinary International Journal*, *2*(01), 37-46.

Nalathambi, D. K., Salleh, K. S. M., Noh, S. H. M., Solaiman, H. S., & Jayaraman, R. (2023). Effort of Politeknik Malaysia as TVET institute in Attaining Sustainable Development Goals (SDGs) Through Twelfth Malaysia Plan. *Borneo Engineering & Advanced Multidisciplinary International Journal*, *2*(01), 37-46.

Naqvi, S. R., Akram, T., Haider, S. A., Khan, W., Kamran, M., Muhammad, N., & Nawaz Qadri, N. (2019). Learning outcomes and assessment methodology: case study of an undergraduate engineering project. *The International Journal of Electrical Engineering & Education*, *56*(2), 140-162.

Nwosu, K. C., Okoyoe, C. C., & Onah, U. H. (2018). An Interpretive Descriptive Study of Factors Affecting Academic Achievement of Underachieving Student Teachers in Nigeria. *Journal of At-Risk Issues*, *21*(2), 20-29.

Popham, W. J. (2018). *Assessment literacy for educators in a hurry*. ASCD.

Pratiwi, D., Arief, M., & Churiyah, M. (2018). The Development of Performance-based Model Authentic Assessment on Archival Subject. *KnE Social Sciences*, 150-166.

Race, P. (2019). *The lecturer's toolkit: A practical guide to assessment, learning and teaching*. Routledge.

Rao, N. J. (2020). Outcome-based education: An outline. *Higher Education for the Future*, *7*(1), 5-21.

Raschwahungu, D. K., Wawire, V., & Kirimi, F. (2023). Strategies for aligning institutional engineering technical vocational education and training practices with industry skills requirements in Kenya. *Reviewed Journal International of Education Practice*, *4*(1), 96-116.

Razalli, N. M., Ahmad, Z. A., Abidin, S. Z. Z., & Musa, H. (2023) Updated Academic Performance Audit (APA) Standards: Implications for Malaysian Polytechnic and Community Colleges.

Ridzuan, M. R., & binti Abd Rahman, N. A. S. (2022). The Analysis of The Government Policy on Technical and Vocational Education and Training (Tvet) and The Predicaments of Tvet In Malaysia. *International Journal of Humanities Technology and Civilization*, 53-58.

Sa, N. T. (2021). The impact of outcome-based philosophy and practice on a group of efl teachers'beliefs and actions. *Journal of science and technology-iuh*, *52*(04).

Saleem, S. M. G., & Gouse, H. S. M. (2019). Comprehensive course-embedded assessment system (CCEAS) to measure program learning outcomes (PLO): Measures that matters. *Health Sciences*, *8*(4), 107-116.

Saleem, S. M. G., & Gouse, H. S. M. (2019). Comprehensive course-embedded assessment system (CCEAS) to measure program learning outcomes (PLO): Measures that matters. *Health Sciences*, *8*(4), 107-116.

Saraya, A., Mardhatillah, A., Fitriani, E. N., & Adiyono, A. (2023). The Role of Islamic Religious Education Teachers: Problems in Evaluating Student Learning Outcomes in Junior High Schools. *Dharmas Education Journal (DE_Journal)*, *4*(2), 565-572.

Schilke, O., Hu, S., & Helfat, C. E. (2018). Quo vadis, dynamic capabilities? A content-analytic review of the current state of knowledge and recommendations for future research. *Academy of management annals*, *12*(1), 390-439.

Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education*, *44*(2), 169-185.

Setiawan, D., Sitorus, J., & Natsir, M. (2018). Inhibiting Factor of Primary School Teacher Competence in Indonesia: Pedagogic and Professionalism. *Asian Social Science*, *14*(6), 30.

Setiawan, D., Sitorus, J., & Natsir, M. (2018). Inhibiting Factor of Primary School Teacher Competence in Indonesia: Pedagogic and Professionalism. *Asian Social Science*, *14*(6), 30.

Shen, X., Wang, X., Zhang, Z., & Fei, L. (2021). Does non-conforming urban development mean the failure of zoning? A framework for conformance-based evaluation. *Environment and Planning B: Urban Analytics and City Science*, *48*(5), 1279-1295.

Suo, Z. (2023). Reform Policy on Internationalization of Higher Education in Malaysia: --Analysis based on Malaysia Education Blueprint 2015-2025 (Higher Education). *Journal of Education and Educational Research*, *5*(1), 154-158.

Tan, K., Chong, M. C., Subramaniam, P., & Wong, L. P. (2018). The effectiveness of outcome-based education on the competencies of nursing students: A systematic review. *Nurse eduation today*, *64*, 180-189.

Tenedero, E. Q., & Pacadaljen, L. M. (2021). Learning experiences in the emerging outcomes-based education (OBE) curriculum of higher education institutions (HEI'S) on the scope of Hammond's evaluation cube. *Psychology and Education*, *9*(58), 2.

Yahia, F. B., Essid, H., & Rebai, S. (2018). Do dropout and environmental factors matter? A directional distance function assessment of tunisian education efficiency. *International Journal of Educational Development*, *60*, 120-127.

Yang, P., Lai, S., Guan, H., & Wang, J. (2022). Teaching Reform and Practice Using the Concept of Outcome-Based Education: A Case Study on Curriculum Design for a Microcontroller Unit Course. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(3), 68-82.

Zaki, N., Turaev, S., Shuaib, K., Krishnan, A., & Mohamed, E. (2023). Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation. *Education and Information Technologies*, 1-20.