Assessment Literacy and Item Development in Higher Education: A Thematic Systematic Review

Zarina Md Ali[1*], Nurfirdawati Muhamad Hanafi[2] ,Noorazman Bin Abd Samad[2]

[1,2]Department of Technology Studies, Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, Malaysia

[1]Department of Civil Engineering, Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn Malaysia, Malaysia

*Corresponding Author

ABSTRACT

Assessment literacy has become a central concern in higher education, particularly as institutions seek to ensure that assessment practices validly and fairly represent student learning. In particular, educators' capacity to develop high-quality assessment items remains a critical yet underexplored dimension of assessment literacy. This review synthesises existing literature on assessment literacy in higher education, with a specific focus on item development practices and related competencies. A thematic systematic literature review was conducted using four academic databases: Scopus, Web of Science, ERIC, and Google Scholar. The review prioritised recent empirical and review studies, while also incorporating seminal works foundational to assessment literacy and cognitive alignment. Relevant studies were selected based on predefined inclusion and exclusion criteria and analysed using thematic synthesis. The findings identify four dominant themes: (1) conceptualizations of assessment literacy; (2) core assessment literacy competencies encompassing knowledge, skills, and attitudes; (3) principles of item development and cognitive alignment; and (4) challenges and gaps in higher education assessment practices. The review highlights a limited availability of structured, practice-oriented frameworks that explicitly support educators in developing cognitively aligned and valid assessment items. Overall, the findings underscore the need for clearer, practice-oriented guidance and targeted professional development initiatives that support educators in translating assessment literacy principles into cognitively aligned assessment items.

INTRODUCTION

Assessment plays a central role in higher education, serving not only as a mechanism for evaluating student learning but also as a key driver of instructional decisions-making and curriculum improvement. Increasingly, the effectiveness of assessment practices is recognised as being closely linked to educators' competence in designing, implementing, and interpreting assessment tasks (Brookhart, 2011; Popham, 2009). As higher education systems place greater emphasis on accountability, learning outcomes, and student-centred approaches, strengthening educators' assessment literacy has become a growing priority.

Assessment literacy is widely conceptualised as a multidimensional construct encompassing educators' knowledge, skills, and attitudes related to assessment practices. Early foundational work by Stiggins (1991, 1995) defined assessment literacy in terms of educators' understanding of assessment principles and their capacity to select and use appropriate assessment methods responsibly. Subsequent scholarship expanded this view by emphasising the importance of professional judgement, ethical considerations, and contextual influences in assessment decision-making (Brookhart, 2011; Xu & Brown, 2016). More recent frameworks further position assessment literacy as a form of professional competence that develops over time through experience, reflection, and engagement with practice (DeLuca et al., 2016; DeLuca et al., 2019).

Within this broader conceptualisation, assessment item development represents a critical yet persistently challenging dimension of assessment literacy in higher education. Assessment items function as the primary tools through which learning outcomes are measured, particularly in examination-oriented and test-based

contexts. The quality of these items is closely linked to their alignment with intended learning outcomes and their capacity to elicit appropriate levels of cognitive engagement (Haladyna et al., 2004; Brookhart, 2018). To support such alignment, cognitive frameworks including Bloom's Taxonomy, the Revised Bloom's Taxonomy, Depth of Knowledge, and the SOLO taxonomy are frequently recommended in assessment design literature (Anderson & Krathwohl, 2001; Webb, 2002; Biggs & Collis, 1982).

Despite the availability of these frameworks, empirical research consistently indicates that many educators encounter difficulties in translating theoretical principles into effective assessment items. Studies report a persistent tendency to design assessment tasks that emphasise lower-order cognitive skills, even when higher-order learning outcomes are explicitly articulated (Brookhart, 2011; DeLuca et al., 2019; Gaikwad et al., 2023). Such misalignment raises concerns regarding the validity and fairness of assessment practices, as well as the extent to which assessments genuinely capture intended learning objectives. These challenges are often compounded by limited formal training in assessment, inconsistent item development practices, and an overreliance on compliance-driven quality assurance mechanisms (DeLuca et al., 2019; Yan & Pastore, 2022).

Although assessment literacy has been widely discussed in educational research, important gaps remain in understanding how assessment competencies which particularly those related to item development are enacted in higher education practice. Existing studies highlight a lack of structured, practice-oriented guidance to support educators in developing cognitively aligned and high-quality assessment items. Consequently, there is a need for a more coherent synthesis of literature that integrates assessment literacy concepts, cognitive alignment frameworks, and principles of assessment item development.

Accordingly, this review synthesises literature on assessment literacy in higher education with a specific focus on item development practices. By examining contemporary empirical and review studies, this thematic systematic literature review aims to identify key conceptual trends, core competencies, and persistent challenges associated with assessment literacy, while also highlighting gaps that may inform future research and professional development initiatives in higher education.

METHODS

Review Design

This study employed a thematic systematic literature review to synthesise existing research on assessment literacy and item development in higher education. A systematic approach was adopted to ensure that literature identification, selection, and synthesis were conducted in a transparent and structured manner. Rather than aggregating statistical findings, the review focused on identifying recurring patterns, concepts, and gaps across studies. The review was conducted through four iterative stages: (1) systematic literature identification across selected databases; (2) screening based on titles, abstracts, and predefined criteria; (3) full-text eligibility assessment; and (4) thematic analysis and synthesis of findings.

Data Sources

The literature search was conducted using four academic databases: Scopus, Web of Science, ERIC, and Google Scholar. These databases were selected to capture peer-reviewed, high-impact studies relevant to assessment literacy, assessment practices, and item development in higher education contexts. The review prioritised recent empirical and review studies, while also including seminal foundational works related to assessment literacy and cognitive alignment to provide theoretical grounding. Consistent with the interpretive nature of a thematic systematic review, principles of the PRISMA framework informed the processes of identification, screening, and selection. However, synthesis emphasised thematic integration rather than numerical reporting.

Search Strategy

A structured keyword search strategy was used to retrieve relevant studies. Search terms were developed based on the review focus and refined through preliminary searches. The primary keywords included *assessment literacy*, *item development*, *assessment design*, *higher education*, and *educators*. Boolean operators such as

"AND" and "OR" were applied to combine keywords and refine search results. The search was limited to studies published in English. Titles and abstracts were initially screened for relevance, followed by full-text screening to confirm alignment with the review objectives. In line with guidance by Xiao and Watson (2017), this review adopted a hybrid approach, selecting methodological elements most appropriate to addressing the research focus rather than adhering rigidly to a single review protocol.

Inclusion and Exclusion Criteria

Clear inclusion and exclusion criteria were applied to ensure consistency and relevance during the screening process. Studies were included if they:

a.   focused on assessment literacy, assessment practices, or item development,
b.   were situated in higher education contexts,
c.   involved educators, lecturers, or academicians as the primary participants; and
d.   were empirical studies or systematic/narrative reviews published in peer-reviewed sources.

Studies were excluded if they:

a.   focused solely on primary or secondary education;
b.   were opinion-based articles, editorials, or commentaries without empirical evidence; or
c.   were not published in English.

Data Analysis

The selected studies were analysed using thematic analysis. Key information was extracted from each study, including research focus, methodological approach, principal findings, and implications related to assessment literacy and item development. Extracted data were compared and grouped based on similarities and recurring issues. Through an iterative process of comparison and refinement, overarching themes were developed to represent dominant trends and gaps in the literature. Although the review followed a sequential process of literature identification, screening, eligibility assessment, and thematic synthesis, numerical reporting at each stage was not emphasised, consistent with the review's thematic and interpretive focus. To enhance transparency and replicability, Table 1 summarises the key stages and procedures of the review process.

Table 1. Summary of Review Protocol and Thematic Analysis Procedures

Stage	Description
Review design	Thematic systematic literature review with an interpretive synthesis focus
Databases	Scopus, Web of Science, ERIC, Google Scholar
Search focus	Assessment literacy, item development, assessment design in higher education
Study selection	Title and abstract screening followed by full-text eligibility assessment
Inclusion criteria	Higher education context; assessment literacy or item development focus; educators as participants
Exclusion criteria	School-level studies; opinion pieces; non-English publications
Time coverage	Emphasis on recent studies, supplemented by seminal foundational works
Analysis approach	Thematic analysis involving iterative coding, comparison, and synthesis
Output	Four interrelated themes capturing concepts, competencies, practices, and gaps

FINDINGS

The thematic analysis of the selected literature identified four interrelated themes that describe how assessment literacy is conceptualised and enacted in higher education, with specific attention to assessment item development. These themes reflect recurring patterns across empirical and review studies rather than isolated findings. Each theme is presented below with supporting evidence from the reviewed literature.

Theme 1: Conceptualisations of Assessment Literacy

Across the reviewed literature, assessment literacy is consistently conceptualised as a multidimensional construct encompassing educators' knowledge, skills, and dispositions related to assessment design, implementation, and interpretation. Table 2 summarises influential conceptualisations of assessment literacy in higher education, highlighting shared dimensions that inform the thematic synthesis.

Table 2. Conceptualizations of Assessment Literacy

Author(s)	Conceptual Emphasis	Key Dimensions Relevant to This Review
Stiggins (1991, 1995)	Assessment literacy as KSA for effective assessment	Knowledge, Skills, Attitudes; validity; reliability; fairness
Brookhart (2011)	Ethical and pedagogical assessment practice	Formative feedback; teacher judgment; student involvement; ethical practice
DeLuca et al. (2019)	Integrated professional competence within systems	Professional judgment; instructional alignment; assessment practice
Popham (2009)	Principles for valid assessment inferences	Validity; reliability; appropriate data interpretation
Xu & Brown (2016)	Contextual and sociocultural aspects of assessment literacy	Teacher agency; institutional context; political awareness
Wiliam (2007, 2011)	Formative assessment to support learning	Feedback mechanisms; continuous improvement; learning focus

Note. Terminology referring to teachers or educators in the original literature is interpreted within the context of higher education academicians in this review.

Early foundational work by Stiggins (1991, 1995) framed assessment literacy primarily in terms of educators' knowledge of assessment principles and their ability to select and apply appropriate assessment methods, with emphasis on validity, reliability, and ethical practice. Subsequent scholarship expanded this view by recognising the complexity of assessment within educational contexts. For example, Xu and Brown (2016) positioned assessment literacy within broader sociocultural and institutional settings, emphasising educators' capacity to make informed assessment decisions responsive to contextual demands.

More recent literature conceptualises assessment literacy as a form of integrated professional competence that develops over time. DeLuca et al. (2016, 2019) describe assessment literacy as an interrelated set of knowledge, skills, and attitudes enabling educators to design, administer, and interpret assessments responsibly. This integrated framing highlights the role of professional judgement, ethical awareness, and reflective practice in assessment-related decision-making.

Across studies, there is broad agreement that assessment literacy extends beyond familiarity with assessment tools to include the ability to design meaningful assessment tasks, interpret assessment evidence accurately, and use assessment outcomes to inform teaching and learning. These conceptualisations establish assessment literacy as a foundational competency for educators in higher education and provide a basis for examining its specific dimensions in relation to item development.

Theme 2: Core Assessment Literacy Competencies: Knowledge, Skills, and Attitudes

Across the reviewed literature, assessment literacy is consistently conceptualised as comprising three interrelated competency domains: knowledge, skills, and attitudes (KSA). This tripartite structure reflects a shared understanding that effective assessment practice in higher education extends beyond technical proficiency to include informed professional judgement and ethical dispositions that shape how assessment is designed, implemented, and interpreted.

Early and influential models frame assessment literacy broadly in terms of educators' general assessment competence across these three domains. As summarised in Table 3, Stiggins' framework emphasises

foundational knowledge of assessment principles, including validity, reliability, and fairness, alongside skills related to assessment creation and interpretation, and professional attitudes grounded in ethics and responsibility. In contrast, Brookhart's model places stronger emphasis on instructionally embedded assessment, highlighting the integration of assessment with teaching and learning processes and the role of teacher judgement in applying assessment principles within classroom contexts. Despite differences in emphasis, both models converge on the importance of aligning assessment items with intended cognitive outcomes and designing assessments that are pedagogically meaningful.

Table 3. Comparison of Key Assessment Literacy Models (Stiggins vs Brookhart)

Aspect	Stiggins (1991, 1995)	Brookhart (2011)	Implications for Item Development
Core orientation	General assessment literacy (KSA)	Instructionally embedded assessment	Need to align item design with learning purpose
Knowledge focus	Validity, reliability, fairness	Alignment with instructional intent	Items must reflect intended cognitive outcomes
Skills emphasis	Assessment creation and interpretation	Practical classroom assessment use	Item construction must be context-sensitive
Attitudinal focus	Ethics and professionalism	Teacher judgment and responsibility	Ethical item design and fair interpretation

Beyond conceptual models, empirical studies highlight how assessment literacy competencies are enacted in practice and the challenges associated with each domain. As synthesised in Table 4, recurring issues related to assessment knowledge include misunderstandings of validity and psychometric principles, superficial use of cognitive frameworks, and overreliance on simplified taxonomies without sufficient attention to cognitive demand. Challenges associated with skills frequently involve difficulties in translating theoretical knowledge into practice, such as aligning items with intended cognitive levels, constructing effective distractors, and providing formative feedback under time and workload constraints. The attitudinal dimension further shapes assessment practice, with several studies reporting tensions between educators' beliefs about assessment for learning and compliance-driven institutional cultures that prioritise accountability over pedagogical development.

Collectively, the findings under Theme 2 indicate that assessment literacy competencies are interdependent rather than discrete. Weaknesses in any single domain, which are knowledge, skills, or attitudes can compromise assessment quality, particularly in item development. The reviewed studies consistently suggest that effective assessment item construction requires not only conceptual understanding of assessment principles but also sustained skill development and professional dispositions that support reflective and context-sensitive assessment practice.

Table 4. Synthesis of Knowledge, Skills, and Attitudes (KSA) Dimensions

Component	Focus	Challenges	Implications	Authors
Knowledge	Conceptual understanding of assessment principles such as validity, reliability, fairness, and alignment.	Misunderstanding of validity and psychometrics; overinflation of cognitive levels; superficial use of frameworks.	Enhance theoretical training in psychometrics, cognitive taxonomies, and framework application.	Messick (1995); Brookhart (2011); Xu & Brown (2016); Chong (2023); Wine & Hoffman (2022); Hamdoun (2022)
Skills	Application of theoretical knowledge through item construction, rubric design, feedback, and data analysis.	Difficulty aligning items with cognitive levels; underuse of formative feedback due to class size, time, and pressure.	Hands-on training in rubric use, digital tools, and adaptive platforms; emphasize personalized feedback.	Brookhart (2011); Mulbar et al. (2017); Ghunaimat (2024); Sulaiman & Ismail (2020); Enu (2021);

| Attitudes | Beliefs and dispositions toward assessment; viewing assessment as a tool for learning. | Gap between beliefs and practice; influence of institutional norms; resistance due to compliance-based culture. | Foster supportive culture; encourage reflective professional development and academician agency. | Brookhart (2011); Tang & Hu (2022); Alonzo et al. (2021); Wine & Hoffman (2022); DeLuca et al. (2013); Xu & Brown (2016) |

Theme 3: Item Development Principles and Cognitive Alignment

Theme 3 focuses on how assessment literacy is operationalised through item development principles and the application of cognitive alignment frameworks. Across the reviewed literature, item development emerges as a central yet persistently challenging component of assessment literacy in higher education. While learning taxonomies and alignment principles are widely referenced in curriculum documents and assessment guidelines, empirical evidence indicates a recurring gap between intended cognitive demand and enacted assessment tasks.

A substantial body of literature identifies constructive alignment as a foundational principle guiding effective item development. Rooted in constructivist learning theory, constructive alignment emphasises coherence between intended learning outcomes, teaching activities, and assessment tasks. As synthesised in Table 5, studies consistently report that alignment supports structured learning experiences, promotes higher-order thinking, and enhances the overall coherence of assessment practices across instructional contexts. However, several studies also caution that alignment is often implemented procedurally, driven by compliance requirements rather than pedagogical intent, which may limit academic autonomy and innovation in assessment design.

Beyond alignment principles, the literature highlights the role of cognitive frameworks in guiding item construction. The Revised Bloom's Taxonomy is most frequently referenced as a practical tool for classifying learning outcomes and assessment items according to cognitive complexity. Complementary frameworks, such as Webb's Depth of Knowledge and the SOLO taxonomy, are also used to support distinctions between surface-level and deeper learning. As summarised in Table 6, each framework offers specific strengths for structuring cognitive demand, while also presenting limitations related to interpretation consistency, usability, and differentiation between adjacent cognitive levels.

Despite the availability of these frameworks, empirical studies consistently document recurring item-writing issues that signal misalignment between learning outcomes and assessment tasks. These issues, synthesised in Table 7, include cognitive misalignment, poor distractor design, technical flaws in item construction, limited item format diversity, and overemphasis on lower-order cognitive skills. Several studies attribute these issues to insufficient professional training in item writing, time constraints, and institutional practices that prioritise procedural compliance over pedagogical quality.

Table 5. Constructive Alignment in Assessment Literacy: Synthesis of Key Contributions in Higher Education Literature

Aspect	Conceptual Synthesis	Evidence across studies	Reported outcomes	Authors
Theoretical Foundation	Grounded in constructivist learning theory; emphasizes alignment of Intended Learning Outcomes (ILOs) with teaching and assessment	Used across disciplines to structure coherent learning experiences.	Promotes deep learning and higher-order thinking.	Biggs (1996); Biggs & Tang (2011)

	practices.			
Core Components	Alignment of learning outcomes, teaching strategies, and assessment tasks to ensure instructional coherence across curricula	Curriculum design, instructional planning, assessment development.	Ensures consistent focus on outcomes across all instructional elements.	Biggs & Tang (2011)
Contemporary Applications	Adapted for digital and AI-supported environments; integrated into project-based and analytics-informed learning platforms.	AI-driven adaptive learning, real-time feedback, learning analytics.	Enhances personalization and learner engagement.	Pereira et al. (2023); Wang et al. (2023); Redecker & Punie (2017)
Gaps / limitations	Risk of superficial implementation driven by compliance rather than pedagogy intent.	Administrative enforcement of alignment practices without pedagogical training.	May limit academic autonomy and hinder innovation.	Loughlin et al. (2020); Hamdoun (2023); Hristov et al. (2023)
Link to Assessment Literacy	Encourages intentional, reflective, and inclusive assessment design aligned with intended learning goals.	Supports item development, instructional alignment, and reflective practice.	Improves coherence and overall quality of assessment practices in higher education.	Biggs & Tang (2011); DeLuca et al. (2019); Yan & Pastore (2022); Pereira et al. (2023)

Table 6. Comparative Frameworks Comparison

Framework/Model	Structure & Focus	Key Strengths in Item Development	Challenges/ Limitations
Bloom's Taxonomy (1956)	Six hierarchical levels: Knowledge to Evaluation; focuses on levels of cognitive complexity	Supports structured item design and alignment with CLOs; promotes higher-order thinking	Assumes linearity; oversimplifies cognitive progression
Revised Bloom's Taxonomy (2001)	Two dimensions: Cognitive Process (e.g., Apply, Create) × Knowledge (e.g., Procedural, Metacognitive)	Enhances clarity and specificity; supports technology-aligned and competency-based assessment	Difficulty distinguishing between higher-order levels; application inconsistency
Depth of Knowledge (Webb, 2002)	Four levels based on thinking complexity: Recall to Extended Thinking	Emphasizes depth of reasoning; aligns tasks with intended learning outcomes	Misconceptions in use; classification may vary with learner proficiency

SOLO Taxonomy (Biggs & Collis, 1982)	Five levels reflecting depth of understanding: Pre-structural to Extended Abstract	Highlights qualitative understanding; supports differentiation and curriculum design	Less commonly used; level distinctions can be subjective

Table 7. Common Item-Writing Issues

Common Issue	Description	Representative Sources
Cognitive Misalignment	Assessment items do not match the cognitive level of intended learning outcomes.	Gaikwad et al. (2023) DeLuca et al. (2019); Haladyna et al., 2002
Poor Distractor Design	Distractors are implausible, grammatically inconsistent, or rarely selected, reducing item effectiveness.	Panadero et al., 2023; Haladyna et al., 2002
Technical Flaws	Errors in item grammar, structure, or format affect clarity, fairness, and reliability.	Haladyna et al., 2002; Brookhart, 2011
Limited Item Format Diversity	Overuse of MCQs or recall-based questions neglects assessment of critical and higher-order thinking.	Gaikwad et al., 2023; Yusoff, 2019
Insufficient Training in Item Writing	Lack of professional development in item writing contributes to recurring flaws in assessment tasks.	Panadero et al., 2023; Gaikwad et al., 2023
Ambiguity and Lack of Clarity	Unclear stems, vague language, or complex sentence structure led to student misinterpretation.	Brookhart, 2011; Haladyna et al., 2002
Bias and Lack of Inclusivity	Assessment items may contain cultural or linguistic bias, disadvantaging certain groups.	DeLuca et al., 2019; Educational Testing Service, 2020
Overemphasis on Lower-Order Skills	Many assessments focus on recall (C1 - C3), limiting cognitive engagement and student thinking.	Brookhart, 2011; Yusoff, 2019

Overall, the findings under Theme 3 indicate that effective item development is not a linear process but an iterative practice involving outcome clarification, cognitive level identification, item drafting, review, and refinement. While emerging studies point to the potential of psychometric analysis and AI-supported tools to enhance item review and cognitive classification, their integration into routine assessment practice remains limited. Collectively, the literature highlights a persistent gap between the conceptual endorsement of cognitive alignment frameworks and their consistent application in assessment item development.

Theme 4: Challenges and Gaps in Higher Education Assessment Practices

Across the reviewed literature, persistent challenges in higher education assessment practices reveal notable gaps in educators' assessment literacy, particularly in relation to item development. One of the most consistently reported issues is the lack of systematic and sustained training in assessment design for higher education educators. Many studies indicate that assessment competencies are often developed informally through teaching experience rather than through structured professional development, resulting in considerable variability and inconsistency in assessment practices across institutions (DeLuca et al., 2019; Yan & Pastore, 2022).

A further prominent challenge concerns the misalignment between intended learning outcomes and assessment items. Despite the widespread adoption of outcome-based education frameworks, educators frequently experience difficulties translating intended learning outcomes into cognitively appropriate assessment tasks. This misalignment is commonly manifested in an overreliance on lower-order cognitive items, even in courses

that explicitly aim to foster higher-order thinking skills (Brookhart, 2011; Gaikwad et al., 2023). Such practices raise concerns regarding the validity, fairness, and interpretability of assessment results, as assessments may fail to accurately capture students' intended learning achievements.

Beyond individual assessment competencies, the literature highlights institutional and contextual constraints that shape assessment practices in higher education. Time pressures, large class sizes, and limited access to assessment support resources are frequently cited as barriers to effective item development. Although quality assurance mechanisms such as moderation and vetting processes are widely implemented, these processes are often perceived as compliance-oriented rather than developmental in nature. As a result, they provide limited guidance for enhancing educators' assessment literacy or improving item quality at the practitioner level.

Importantly, the reviewed studies point to a significant gap in the availability of structured, practice-oriented frameworks that explicitly guide educators in developing high-quality assessment items. While many studies advocate strengthening assessment literacy, relatively few offer concrete models that integrate assessment principles, cognitive alignment frameworks, and practical item-writing strategies. This lack of operational guidance makes it challenging for educators to translate theoretical knowledge into effective assessment practice, particularly within examination-oriented higher education contexts.

Across the literature, a consistent disconnect emerges between conceptual discussions of assessment literacy and its practical enactment in assessment item development. Although assessment literacy frameworks frequently articulate broad competencies and principles, limited attention is given to the systematic development of item-writing skills that align cognitive frameworks with real assessment tasks. Existing research often prioritises policy-level assessment principles or institutional implementation, leaving item-level practices under-theorised and under-supported.

To synthesise these recurring challenges, Table 8 summarises key areas where assessment literacy research and practice remain insufficiently developed. Collectively, the findings indicate that assessment item development represents a critical yet under-addressed dimension of assessment literacy in higher education, highlighting the need for more coherent, context-sensitive approaches that bridge assessment theory and day-to-day assessment practice.

Table 8. Key Gaps Identified in Assessment Literacy Literature

Area	What Is Well Established	Identified Gap
Assessment literacy	Conceptual definitions and competencies	Limited item-level operational guidance
Cognitive frameworks	Bloom, DoK, SOLO widely referenced	Weak translation into item construction
Professional development	General assessment training	Lack of structured item-development guidelines
Higher education context	Policy-driven assessment practices	Limited support for academicians' assessment literacy

These gaps underscore the need for structured, practice-oriented frameworks that explicitly support educators in translating assessment literacy principles into high-quality assessment items.

DISCUSSION

This thematic systematic review demonstrates that assessment literacy in higher education is consistently conceptualised as an integrated professional competence, encompassing educators' knowledge, skills, and attitudes rather than a narrowly defined technical capability. Across diverse higher education contexts, the consistency of findings suggests that the identified themes represent systemic and recurring patterns rather than isolated challenges. Contemporary literature positions assessment literacy as dynamic and developmental, shaped by professional judgement, ethical awareness, reflective practice, and engagement with institutional contexts.

Despite this broadened conceptualisation, the review reveals a persistent disconnect between educators' theoretical understanding of assessment and its enactment in everyday assessment practices. A central contribution of this review lies in foregrounding item development as a critical yet under-examined dimension of assessment literacy. Although cognitive alignment frameworks—such as Revised Bloom's Taxonomy, Depth of Knowledge, and SOLO—are widely referenced in curriculum documents and assessment guidelines, many educators continue to struggle to operationalise these frameworks during item construction. Empirical evidence consistently reports an overreliance on lower-order cognitive items, even in courses explicitly designed to assess higher-order learning outcomes. This misalignment highlights a gap between conceptual knowledge of cognitive frameworks and their practical application, raising ongoing concerns regarding the validity, fairness, and interpretability of assessment outcomes.

Figure 1 synthesises these findings into a practice-oriented framework illustrating how assessment literacy competencies (knowledge, skills, and attitudes) inform item development through the application of cognitive alignment frameworks. Importantly, the framework situates item development within broader institutional and professional contexts, recognising that factors such as professional development opportunities, assessment policies, and assessment culture can either enable or constrain educators' item development competence. This contextual dimension helps explain why technically sound assessment principles are not always consistently reflected in practice.
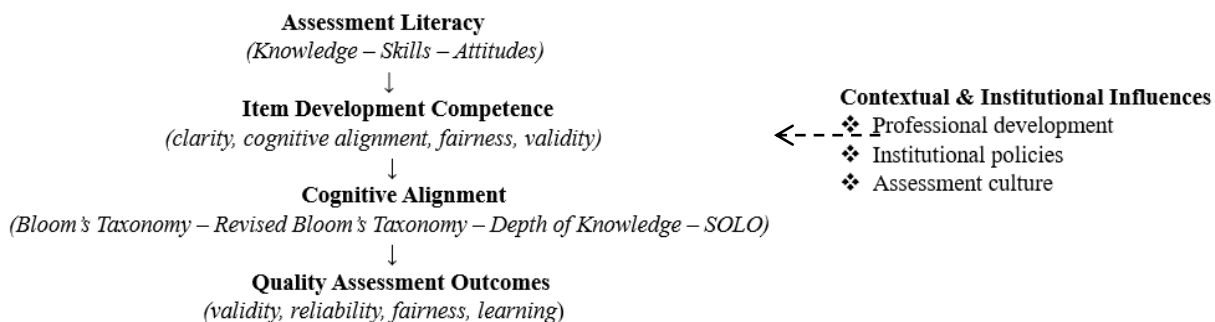

Assessment Literacy
(Knowledge – Skills – Attitudes)
↓
Item Development Competence
(clarity, cognitive alignment, fairness, validity)
↓
Cognitive Alignment
(Bloom's Taxonomy – Revised Bloom's Taxonomy – Depth of Knowledge – SOLO)
↓
Quality Assessment Outcomes
(validity, reliability, fairness, learning)

Contextual & Institutional Influences
❖ Professional development
❖ Institutional policies
❖ Assessment culture

Figure 1. A practice-oriented conceptual framework showing how assessment literacy competencies inform assessment item development through cognitive alignment, within institutional and professional contexts, to enhance assessment quality in higher education.

Note. Institutional and contextual influences, including professional development opportunities, assessment policies, and assessment culture, may function as enabling or constraining factors shaping educators' item development competence.

How to Use the Framework in Practice

1.   Start from the Course Learning Outcomes (CLOs) and the intended cognitive demand.

2.   Select an appropriate cognitive framework (Revised Bloom's Taxonomy, Depth of Knowledge, or SOLO) and map the assessment item accordingly.

3.   Apply item-writing principles, including clarity of wording, fairness, and the construction of effective distractors.

4.   Conduct review and validation through moderation processes, supported by classical test theory or psychometric checks where feasible.

The review further highlights the influence of systemic and institutional constraints on assessment practices. Limited access to structured professional development, time pressures, large class sizes, and a strong emphasis on compliance-driven quality assurance processes are commonly reported barriers. While moderation and

vetting mechanisms are intended to safeguard assessment quality, they often prioritise procedural correctness over developmental feedback. As a result, opportunities for formative feedback, reflective practice, and capacity building in item development remain underutilised.

Importantly, the findings point to a notable gap in the availability of practice-oriented frameworks or guidelines that explicitly support educators in developing high-quality assessment items. While many studies advocate strengthening assessment literacy, relatively few offer concrete, actionable models that integrate assessment principles, cognitive alignment, and item-writing strategies in a coherent manner. This gap underscores the need for structured and context-sensitive approaches that bridge theoretical understanding and day-to-day assessment practice in higher education.

Practical Implications: A Practice-Oriented Framework for Item Development

The findings of this review have clear practical implications for assessment practice in higher education, particularly in relation to item development. The proposed framework is intended for use by lecturers, examination committee members, and assessment moderators who are directly involved in designing, reviewing, and approving assessment items. It can be applied across key assessment stages, including item writing, assessment blueprinting, and moderation or vetting processes.

When applied during item construction, the framework supports educators in aligning assessment items with intended learning outcomes and appropriate cognitive levels by integrating assessment literacy competencies with established cognitive alignment frameworks. During moderation and review, it provides a structured lens for evaluating item clarity, fairness, and cognitive demand, helping to identify common technical flaws and misalignment issues. Overall, applying this framework is expected to result in more coherent assessment design, clearer cognitive targeting, and a reduction in flawed or superficial items, thereby enhancing the validity, reliability, and fairness of assessment practices in higher education.

Key Insights from the Review

Drawing together the evidence across themes, this review highlights several consistent insights that characterise current assessment literacy practices in higher education:

• Item development is a central expression of assessment literacy, rather than a peripheral or purely technical task.

• Cognitive frameworks are widely referenced but remain difficult to operationalise at the item level.

• Misalignment between intended learning outcomes and assessment items commonly leads to an overuse of lower-order cognitive questions.

• Institutions require integrated, practice-oriented guidance that links cognitive alignment, item-writing principles, and validation processes to support consistent and high-quality assessment practices.

CONCLUSION

This systematic review synthesised contemporary literature on assessment literacy in higher education, with a particular focus on assessment item development as a core component of effective assessment practice. Through thematic synthesis, the review identified key trends in how assessment literacy is conceptualised, the core competencies expected of educators, and the recurring challenges associated with designing cognitively aligned assessment items. Overall, the findings reaffirm that assessment literacy extends beyond technical knowledge to encompass practical skills and professional dispositions that collectively shape assessment quality. A central insight from this review is that item development remains a persistent area of concern in higher education assessment. Despite the widespread adoption of outcome-based education and cognitive frameworks, many educators continue to experience difficulties in translating intended learning outcomes into valid and cognitively appropriate assessment items. This misalignment highlights an enduring gap between theoretical understanding and practical application, particularly in the area of assessment item construction.

Importantly, the review reveals a limited availability of structured, practice-oriented frameworks that explicitly guide educators in developing high-quality assessment items. While existing studies consistently emphasise the importance of strengthening assessment literacy, fewer provide concrete and coherent models that integrate assessment principles, cognitive alignment, and item-writing strategies. Addressing this gap requires greater emphasis on targeted professional development initiatives that support educators in refining their assessment competencies and engaging in reflective assessment practice. In conclusion, strengthening assessment literacy, particularly in relation to item development, which has significant implications for the validity, reliability, and fairness of assessment in higher education. Future research should prioritise the development and evaluation of structured assessment literacy frameworks that assist educators in translating assessment theory into effective practice. Such efforts have the potential to enhance assessment quality and contribute to improved student learning outcomes across diverse higher education contexts. By foregrounding item development as a core enactment of assessment literacy, this review shifts attention from policy-level discourse to the practical realities of assessment design in higher education.

ACKNOWLEDGEMENTS

REFERENCES

1. Alonzo, D., Liu, Y., & Weng, W. (2021). Assessing Teachers' Assessment Literacy: A Review of the Literature. Educational Measurement: Issues and Practice, 40(1), 1–12.
2. Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
3. Biggs, J. B. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364. https://doi.org/10.1007/BF00138871
4. Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy. Academic Press.
5. Biggs & Tang (2011) Biggs, J., & Tang, C. (2011). Teaching for quality learning at university: What the student does. 4th ed. McGraw-Hill Education.
6. Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x
7. Brookhart, S. M. (2018). How to create and use rubrics for formative assessment and grading. ASCD.
8. Chong, W. H. (2023). Validation of AI-driven assessment tools for enhancing teacher assessment literacy. Journal of Educational Technology and Development, 14(2), 180-198.
9. DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment literacy. Teachers and Teaching, 19(4), 430–443. https://doi.org/10.1080/13540602.2013.827366
10. DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. Educational Assessment, Evaluation and Accountability, 28(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6
11. DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., & Gibson, A. (2019). Policies, programs, and practices: Exploring the complex dynamics of assessment education. Assessment in Education: Principles, Policy & Practice, 26(2), 1–22. https://doi.org/10.1080/0969594X.2019.1571994
12. Educational Testing Service. (2020). The Nation's Report Card: Reading 2019. Washington, DC: National Center for Education Statistics.
13. Enu, J. A. (2021). Assessment Literacy among Ghanaian Teachers: Implications for Educational Practice. International Journal of Educational Research, 105, 101713.
14. Gaikwad, S., Chavan, R., & Patil, S. (2023). Challenges in assessing higher-order thinking skills in higher education. Journal of Education and Learning, 12(2), 45–55.
15. Ghunaimat, M. (2024). Enhancing Assessment Literacy among Jordanian Teachers: Challenges and Opportunities. Middle Eastern Journal of Educational Research, 12(2), 89–102.

16. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309-334.

17. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2004). Constructing written test questions for the basic and clinical sciences. NBME.

18. A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 17(3), 309–334. https://doi.org/10.1207/s15324818ame1703_5

19. Hamdoun, W. M. A. (2023). Constructive alignment approach: Enhancing learning and teaching. British Journal of Multidisciplinary and Advanced Studies, 4(2), 162–170.

20. Hristov, S., Nakov, D., & Miočinović, J. (2023). Constructive alignment between objectives, teaching and learning activities, student competencies and assessment methods in higher education. Journal of Agriculture and Plant Sciences, 21(2), 21–36. https://doi.org/10.46763/JAPS23212021h

21. Loughlin, C., van der Kleij, F., & Wyatt-Smith, C. (2020). Teacher professional learning to support classroom assessment: An evidence review. Australian Educational Researcher, 47(3), 491–514. https://doi.org/10.1007/s13384-019-00347-7

22. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

23. Mulbar, U., Rahman, M. A., & Ahmar, A. S. (2018). Analysis of the students' ability in solving mathematics problems based on SOLO Taxonomy. Advances in Social Science, Education and Humanities Research, 160, 52–56.

24. Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. Educational Psychology Review, 35, 113. https://doi.org/10.1007/s10648-023-09823-4

25. Pereira, D., Flores, M. A., & Niklasson, L. (2023). Assessment cultures and learning cultures: Exploring the interplay. Assessment & Evaluation in Higher Education, 48(2), 1–15. https://doi.org/10.1080/02602938.2022.2123456

26. Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? Theory Into Practice, 48(1), 4–11. https://doi.org/10.1080/00405840802577536

27. Redecker, C., & Punie, Y. (2017). European framework for the digital competence of educators: DigCompEdu. Publications Office of the European Union. https://doi.org/10.2760/159770

28. Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72(7), 534–539.

29. Stiggins, R. J. (1995). Student-centered classroom assessment. Merrill.

30. Sulaiman, T., & Ismail, H. (2020). Teachers' perceptions of 21st-century assessment practices. Malaysian Journal of Learning and Instruction, 17(2), 39-62.

31. Tang, S., & Hu, Y. (2022). Exploring the relationship between assessment literacy and teaching efficacy among EFL teachers in China. Language Assessment Quarterly, 19(1), 1–20.

32. Wang, X., Liu, C., & Zhang, Y. (2023). Integrating Constructive Alignment into AI-driven adaptive learning platforms. Educational Technology Research and Development, 71(1), 65-82. https://doi.org/10.1007/s11423-022-10123-6

33. Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. Wisconsin Center for Education Research.

34. Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), Second handbook of mathematics teaching and learning (pp. 1053–1098). Information Age Publishing.

35. Wiliam, D. (2011). Embedded formative assessment. Solution Tree Press.

36. Wine, D., & Hoffman, J. (2022). Strengthening Depth of Knowledge in assessment practice. Educational Practice Review, 20(4), 35-50.

37. Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education, 58, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

38. Yan, Z., & Pastore, S. (2022). Assessment literacy for educators: Theory and practice. Educational Assessment, Evaluation and Accountability, 34(2), 1–20. https://doi.org/10.1007/s11092-021-09362-0

39. Yusoff, N. M. (2019). Assessment literacy among Malaysian secondary school teachers: A case study. Malaysian Journal of Learning and Instruction, 16(2), 1–20.