

Relational Validity in Practice: A Decade of Research on Alternative Assessment in English Language Teaching Higher Education (2014–2024)

Arathai Din Eak*, Li Hsien Ooi, Saw Fen Tan

School of Education, Humanities and Social Sciences, Wawasan Open University, Malaysia

*Corresponding author: arathaide@wou.edu.my

Received: 25 August 2025; **Revised:** 5 December 2025;
Accepted: 19 December 2025; **Published:** 27 December 2025

To link to this article: <https://doi.org/10.37134/ajatel.vol15.2.7.2025>

Abstract

The past decade has witnessed a marked shift in English Language Teaching (ELT) assessment within higher education from standardised testing toward formative and learner-centred approaches. This study maps and reinterprets this evolution through a bibliometric and AI-assisted thematic analysis of 516 peer-reviewed journal articles published between 2014 and 2024. Using *Dimensions.ai* as the data source, co-word mapping and Latent Dirichlet Allocation (LDA) modelling identified five thematic clusters representing teacher assessment literacy, learner-centred experimentation, systemic reform, inclusivity, and sociocultural feedback dynamics. Although publication output increased exponentially after 2019, conceptual integration between sociocultural learning theory and validation frameworks remains limited. To address this gap, the study introduces relational validity as a synthesising framework that aligns evidential reasoning with dialogic, ethical, and pedagogical dimensions of assessment. This construct reframes validity as coherence among evidence, interpretation, and educational consequence, extending traditional psychometric concerns toward a more contextually grounded, human-centred paradigm. By combining structural bibliometric analysis with interpretive theorisation, this study offers both an overview of alternative assessment research and a conceptual roadmap for advancing fairness, reliability, and transformative practice in ELT higher education.

Keywords: Alternative Assessment, Bibliometric Analysis, Relational Validity, English Language Teaching, Higher Education

INTRODUCTION

Assessment in English Language Teaching (ELT) higher education has undergone a significant pedagogical reorientation over the past decade, with increasing scholarly attention directed toward approaches that move beyond standardised, summative testing and toward more formative and learner-centred practices (Black & Wiliam, 2010; Earl, 2012; Shepard, 2019). Traditionally, assessment in ELT was dominated by large-scale, test-based models designed to measure discrete aspects of linguistic competence and to support high-stakes decision-making. More recent scholarship, however, has increasingly conceptualised assessment as an integral component of learning, emphasising feedback, reflection, and learner participation as central mechanisms for supporting development in higher education contexts.

Within this evolving landscape, alternative assessment has emerged as an important area of inquiry. The term is commonly used to describe assessment approaches that depart from conventional, time-bound testing formats and instead prioritise sustained evidence of learning, learner engagement, and contextualised judgement. Practices frequently associated with alternative assessment include portfolios, peer assessment, self-assessment, and project-based tasks (Brown & Hudson, 1998; Carless,

2015). While these approaches are often discussed alongside formative, authentic, and performance-based assessment, the boundaries between these constructs are not always consistently articulated in the literature. Formative assessment is typically defined by its function, namely the provision of feedback to support subsequent learning (Cheng & Fox, 2017). Authentic assessment foregrounds tasks that approximate real-world language use beyond the classroom (Palm, 2008), while performance-based assessment focuses on learners' demonstration of knowledge and skills through specified tasks under defined conditions (Panadero et al., 2016). In this study, alternative assessment is used as an umbrella term to refer to assessment practices that actively involve learners in evaluative processes and that emphasise learning development across time rather than isolated performance outcomes.

A growing body of ELT research reports pedagogical benefits associated with alternative assessment practices, including increased learner engagement, enhanced awareness of criteria, and expanded opportunities for reflection and self-regulation (Rea-Dickins, 2004; Cheng & Fox, 2017; Andrade & Brookhart, 2016). These outcomes are particularly salient in higher education settings, where assessment is expected to support extended learning trajectories and the development of transferable academic and professional competencies. At the same time, the literature remains heterogeneous in both theoretical orientation and methodological design, contributing to uneven conceptual development and variable interpretations of assessment quality.

A central challenge underpinning this body of work concerns the theoretical framing of assessment quality. Ongoing debates reflect differing views on whether assessment should be understood primarily as an evidential enterprise grounded in psychometric principles such as reliability and validity (Fulcher & Davidson, 2007), or as a socially mediated practice shaped by interaction, context, and learner identity (Scarino, 2013; Shohamy, 2020). These positions are rooted in two influential but often separately developed traditions. Validation theory emphasises the justification of assessment interpretations and decisions through systematic evidential reasoning, with attention to fairness and educational consequence (Messick, 1996; Kane, 2013). Sociocultural learning theory, by contrast, conceptualises assessment as a dialogic and mediated process embedded within learning activity, where meaning and competence are co-constructed through participation and feedback (Vygotsky, 1978; Lantolf & Poehner, 2011).

Recent scholarship has increasingly highlighted the difficulty of applying traditional validity frameworks to classroom-based and alternative assessment contexts, where evidence is longitudinal, interactional, and shaped by pedagogical relationships (Phakiti & Isaacs, 2021; Taylor & Banerjee, 2023). While validation-oriented studies often privilege technical consistency and inferential control, they may underrepresent the interpretive and ethical dimensions of assessment practice. Conversely, sociocultural research foregrounds mediation, learner agency, and identity formation but frequently provides limited articulation of how assessment interpretations are warranted beyond local contexts. As a result, alternative assessment research in ELT higher education has developed in ways that are empirically rich yet conceptually segmented.

To address this theoretical separation, the present study introduces the concept of *relational validity* as a synthesising interpretive framework. Building on established models of validity (Messick, 1989; Kane, 2013) and sociocultural perspectives that conceptualise assessment as mediated and consequential practice (Scarino, 2013; Shepard, 2019), relational validity frames assessment quality as coherence among three interdependent dimensions: evidence, interpretation, and educational consequence. Rather than positioning validity as a procedural checklist, this framework emphasises the situated justification of assessment decisions and the ethical implications of assessment practices within higher education environments.

This study adopts a bibliometric approach to examine how alternative assessment research in ELT higher education has evolved between 2014 and 2024. By analysing 516 peer-reviewed journal articles using co-word mapping and AI-assisted thematic analysis, the study identifies publication trends, dominant thematic clusters, and patterns of conceptual convergence and divergence across the field. Through this dual analytical lens, the study seeks to provide both an empirical overview of alternative assessment research and a theoretically informed basis for advancing more coherent, contextually grounded approaches to assessment in ELT higher education.

LITERATURE REVIEW

Assessment in English Language Teaching (ELT) higher education has evolved substantially over recent decades, moving from traditions grounded in measurement and standardisation toward approaches that foreground learning processes, reflection, and classroom interaction. Earlier assessment scholarship largely conceptualised assessment as a measurement activity, with primary concerns centred on reliability, objectivity, and comparability of scores for decision-making purposes (Messick, 1996; Shepard, 2000). Within this paradigm, language ability was often treated as a stable construct that could be sampled through controlled tasks and represented through numerical outcomes. While such approaches offered procedural clarity and administrative efficiency, they have been widely critiqued for underrepresenting the situated, adaptive, and socially contingent nature of language use, particularly in higher education contexts where learning unfolds through extended participation in disciplinary practices (Brown & Hudson, 1998; Lynch, 2001).

In response to these limitations, alternative assessment has gained increasing prominence in ELT higher education, particularly through practices such as portfolios, reflective journals, peer assessment, self-assessment, and project- or task-based assessment (Brown & Hudson, 1998; Carless, 2015). Although alternative assessment is frequently discussed alongside related constructs, the boundaries between these approaches are not always clearly defined in the literature. Formative assessment is primarily characterised by its function, namely the generation of feedback intended to inform subsequent teaching and learning (Cheng & Fox, 2017). Authentic assessment emphasises tasks that approximate real-world or professional language use beyond the classroom (Palm, 2008). Performance-based assessment focuses on learners' demonstration of knowledge and skills through specified tasks assessed against predefined criteria, often within structured conditions (Panadero et al., 2016). In this study, alternative assessment is conceptualised as an umbrella category that foregrounds learner engagement in evaluative processes, iterative evidence collection, and the integration of assessment for learning and assessment as learning, positioning learners as active contributors to judgement rather than passive recipients of scores (Andrade & Brookhart, 2016).

Empirical studies across higher education contexts frequently associate alternative assessment practices with pedagogical benefits such as increased learner engagement, enhanced metacognitive awareness, and deeper interaction with assessment criteria (Rea-Dickins, 2004; Cheng & Fox, 2017). These benefits are particularly salient in tertiary settings, where assessment is expected to support sustained language development and the acquisition of transferable academic and professional competencies. At the same time, the literature remains heterogeneous in both conceptual framing and methodological approach. Key terms are sometimes used interchangeably, and studies may prioritise implementation outcomes without explicitly situating practices within coherent assessment theory, contributing to conceptual ambiguity and uneven theoretical development (Palm, 2008; Panadero et al., 2016).

A persistent source of fragmentation in alternative assessment research concerns differing epistemological orientations toward assessment quality. One dominant orientation is measurement-oriented, conceptualising assessment quality primarily in terms of technical defensibility, including reliability, standardisation, and evidence-based inference. A second orientation is mediation-oriented, viewing assessment as embedded within social practice and emphasising interaction, learner agency, and contextual meaning-making. This division is particularly salient in language assessment, where performance is inherently contextual and dynamic, yet assessment decisions often require justification beyond individual classroom interactions.

Sociocultural theory conceptualises learning as mediated activity shaped through interaction, cultural tools, and participation in communities of practice (Vygotsky, 1978). From this perspective, assessment is not merely an endpoint judgement but an integral component of learning itself, enacted through feedback dialogue, peer interaction, self-assessment, and teacher mediation (Scarino, 2013; Carless & Boud, 2018). Assessment practices are understood as relational, shaping learner identity, positioning students within academic communities, and supporting the development of self-regulation and evaluative judgement. Alternative assessment approaches align closely with sociocultural assumptions insofar as they create opportunities for dialogue about quality, iterative improvement, and shared engagement with criteria.

At the same time, sociocultural research in ELT assessment has been critiqued for limited engagement with explicit validation procedures. Studies are often qualitative and context-specific, offering rich accounts of classroom practice and learner experience while providing fewer systematic warrants for the defensibility of interpretations beyond local settings (Davison & Leung, 2009; Phongsirikul, 2018). As a result, fairness and validity may be discussed primarily in experiential or ethical terms rather than articulated through structured inferential arguments supported by multiple forms of evidence.

Validity theory provides the dominant framework for justifying assessment interpretations and decisions. Messick's unified validity framework conceptualises validity as an integrated evaluative judgement supported by evidence relating to construct representation, relevance, and the social consequences of score use (Messick, 1989, 1996). Kane's argument-based approach further operationalises validity as the evaluation of an interpretive argument, requiring explicit articulation of claims, warrants, and rebuttals across stages such as scoring, generalisation, explanation, and decision-making (Kane, 2013). Within these frameworks, fairness is treated as an integral component of validity rather than as a separate ethical consideration.

However, in ELT higher education contexts, validity is often operationalised in relatively narrow terms, frequently equated with reliability or internal consistency, particularly where assessments are locally designed and resource constraints limit large-scale validation efforts (Taylor & Banerjee, 2023). Moreover, the application of traditional validity frameworks to classroom-based and alternative assessment contexts remains contested. Alternative assessment generates evidence that is longitudinal, dialogic, and shaped by pedagogical relationships, characteristics that challenge conventional assumptions about standardisation and comparability (Phakiti & Isaacs, 2021). Consequently, validation theory and sociocultural theory often develop in parallel rather than in productive integration within the literature.

To address this theoretical separation, the present study adopts *relational validity* as an integrative construct for interpreting assessment quality in alternative assessment research. Relational validity conceptualises validity as coherence across three interdependent dimensions: empirical evidence, contextual interpretation, and educational consequence. Drawing on established validity theory (Messick, 1989; Kane, 2013) while incorporating sociocultural commitments to mediation, dialogue, and ethical participation (Scarino, 2013; Shepard, 2019), this framework reframes reliability as contextual consistency supported through transparent criteria and shared understanding, rather than solely as statistical uniformity. Fairness is conceptualised as an emergent property of participatory assessment cultures, where learners have access to standards, opportunities to develop evaluative judgement, and meaningful avenues to understand and contest assessment decisions (Kunnan, 2017; Shohamy, 2020).

Although debates concerning assessment quality are well established, it remains unclear how alternative assessment research in ELT higher education has integrated measurement-oriented validity arguments with sociocultural accounts of mediation, identity, and participation. Bibliometric analysis provides a systematic means of examining this issue at scale by mapping publication trends, conceptual structures, and thematic development over time. Previous bibliometric studies in related domains have demonstrated the capacity of such methods to reveal intellectual trajectories and thematic concentrations within language assessment research (Altinay et al., 2024; Lin & Yu, 2023; Zhang, 2025). Building on this work, the present study applies co-word mapping and AI-assisted topic modelling to examine how concepts associated with validity theory and sociocultural perspectives co-occur, cluster, and evolve within alternative assessment research in ELT higher education.

RESEARCH DESIGN

This study adopts a mixed-method bibliometric research design that integrates co-word mapping with AI-assisted thematic analysis to examine the conceptual evolution of alternative assessment research in English Language Teaching (ELT) higher education between 2014 and 2024. The design is explanatory and exploratory in nature. It aims not only to document publication growth but also to analyse how key theoretical constructs, including validity, fairness, reliability, and educational consequence, are articulated and interconnected across the literature.

The integration of bibliometric techniques and AI-assisted thematic analysis follows a sequential and complementary logic. Bibliometric analysis provides a structural overview of the field by identifying patterns of keyword co-occurrence and thematic clustering across a large corpus. This quantitative mapping is subsequently supplemented by AI-assisted thematic analysis, which supports interpretive examination of latent semantic patterns within the same dataset. Together, these methods enable a systematic investigation of both the surface structure and the underlying conceptual organisation of alternative assessment research.

1. Bibliometric Analysis

The first phase of analysis involved co-word mapping conducted using VOSviewer (version 1.6.20; van Eck & Waltman, 2010). Co-word analysis focuses on the frequency and co-occurrence of key terms within titles and abstracts, operating on the assumption that frequently co-occurring terms reflect shared conceptual space within a research field. Titles and abstracts were selected as the unit of analysis because they encapsulate authors' central conceptual framing while allowing consistent comparison across a large corpus.

Noun phrases were automatically extracted and subsequently normalised to address lexical variation and synonymy. For example, variants such as "e-portfolio" and "electronic portfolio" were consolidated under a single term. Terms with low frequency were excluded to reduce noise and enhance the interpretability of the resulting network. The Louvain clustering algorithm was applied with a resolution parameter of 1.0 to detect thematic clusters based on co-occurrence strength. The resulting network visualisation represents thematic proximity rather than causal or hierarchical relationships and is interpreted accordingly.

2. AI-Assisted Thematic Analysis

The second phase of analysis employed AI-assisted topic modelling to explore latent thematic structures within the corpus. A natural language processing pipeline implemented in Python was used for data preprocessing, including tokenisation, lemmatisation, and stopword removal. Texts were vectorised using term frequency-inverse document frequency (TF-IDF) weighting to balance term salience across documents.

Latent Dirichlet Allocation (LDA) was then applied to identify probabilistic topic distributions across the corpus. A model with 20 topics was selected based on coherence optimisation ($CV = 0.62$) and iterative refinement across 100 iterations. The purpose of LDA in this study is not to produce definitive thematic labels but to surface recurring semantic patterns that complement the structural insights generated through co-word analysis.

To enhance interpretive validity, each LDA-derived topic was systematically compared with its corresponding bibliometric cluster. This cross-validation process ensured alignment between the quantitative network structure and the semantic content of topics. AI-generated topic labels were subsequently reviewed by two independent experts in language assessment. Inter-rater reliability ($IRR = 0.87$) indicates a high level of agreement in thematic interpretation. Discrepancies were resolved through discussion to reach consensus, strengthening the interpretive coherence of the findings.

3. Analytical Orientation and Validity Considerations

The methodological design of this study is informed by the principle of relational validity. In this context, validity is not treated solely as a statistical property but as coherence between empirical evidence, interpretive reasoning, and analytical consequence. Bibliometric indicators provide empirical structure, AI-assisted modelling supports interpretive depth, and expert validation ensures disciplinary grounding. Claims derived from the analysis are therefore framed as field-level patterns and tendencies rather than as causal or predictive assertions.

It is acknowledged that bibliometric and topic-modelling techniques operate on abstracted textual representations and do not capture the full methodological or contextual detail of individual studies. Accordingly, interpretations focus on conceptual trends, thematic prominence, and patterns of association rather than evaluative judgments of study quality.

DATA SOURCE AND CORPUS SELECTION

The Dimensions.ai database was selected as the primary data source due to its extensive coverage of peer-reviewed journals in education and applied linguistics. Compared to Scopus and Web of Science, Dimensions.ai offers inclusivity of emerging and non-indexed ELT research, particularly from regional and open-access journals (Hook et al., 2018; Visser et al., 2021). The database's ability to capture a broader diversity of English Language Teaching (ELT) research is vital for assessing inclusivity and equity trends in alternative assessment. However, it is acknowledged that Dimensions.ai may have limitations in representing non-English publications or research from non-indexed journals. This selection is justified as it enhances the comprehensiveness of the study's conceptual analysis and mirrors global trends in the field.

The search was conducted across titles and abstracts using the following Boolean query as presented in Table 1.

Table 1 Search strategy and selection criteria

Category	Parameters
Database	Dimensions.ai
Search Terms	("alternative assessment" OR "authentic assessment" OR "performance-based assessment" OR "formative assessment" OR "portfolio assessment" OR "peer assessment" OR "self-assessment") AND ("English language teaching" OR "ELT" OR "TESOL" OR "ESL" OR "EFL") AND ("higher education" OR "university" OR "college")
Fields	Title, Abstract
Language	English
Document Type	Peer-reviewed journal articles
Timeframe	2014–2024
Inclusion Criteria	Empirical or conceptual studies focused on alternative assessments in ELT within higher education contexts
Exclusion Criteria	Grey literature, K–12 studies, non-ELT domains, editorials, book chapters, and duplicates
Final Corpus	516 articles

The search was restricted to English-language, peer-reviewed journal articles published between 2014 and 2024. Excluded were editorials, conference papers, book chapters, and K–12 studies. Following PRISMA (2020) procedures (Page et al., 2021), duplicates were removed and abstracts screened for relevance, yielding a final corpus of 516 publications. Figure 1 presents the PRISMA flow diagram of the identification and screening process.

This study adhered to ethical guidelines set forth by the AERA (2011) Code of Ethics, emphasising transparency, replicability, and responsible data handling. Since all data are drawn from publicly available metadata and do not involve human participants, ethical concerns related to participant consent or privacy are not applicable.

However, several limitations are acknowledged. First, reliance on Dimensions.ai may underrepresent non-English publications and research published in non-indexed venues. Second, analysis at the title and abstract level constrains access to full methodological detail. Third, topic modelling relies on probabilistic approximation, which may introduce interpretive uncertainty. These limitations were mitigated through expert validation and triangulation between bibliometric and AI-assisted analyses.

Despite these constraints, the research design offers a transparent and replicable approach to mapping conceptual development within alternative assessment research. By integrating structural mapping with interpretive analysis, the study provides a robust methodological foundation for examining how validity-oriented and sociocultural perspectives have evolved and interacted within ELT higher education scholarship.

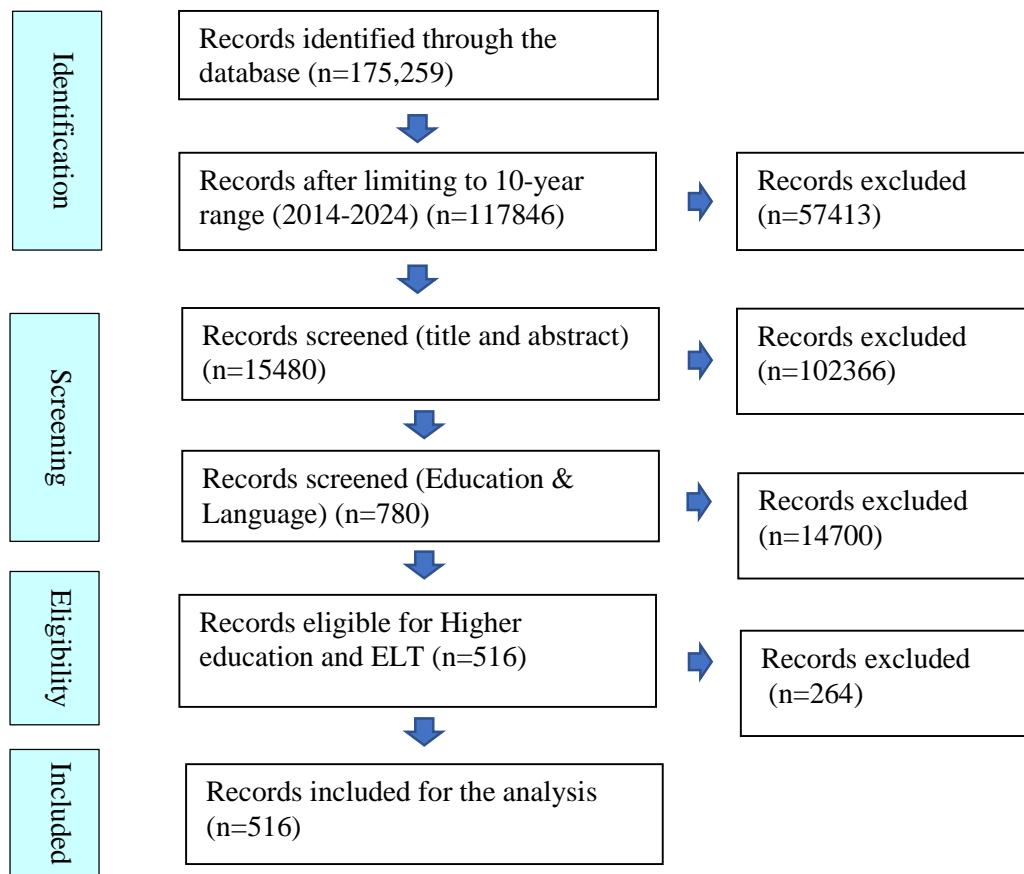


Figure 1 The PRISMA diagram

RESULTS

1. Publication Trends (2014–2024)

The bibliometric analysis identified a total of 516 peer-reviewed journal articles on alternative assessment in English Language Teaching (ELT) higher education published between 2014 and 2024. As shown in Figure 2, publication output increased steadily over the period, rising from 17 articles in 2014 to 104 articles in 2024. This represents a cumulative growth rate of approximately 512 percent, with a strong linear fit ($R^2 = .93$), indicating sustained expansion of scholarly attention to alternative assessment in tertiary ELT contexts.

Based on visual inspection of annual publication frequencies, three analytically distinct phases can be identified. These phases are not treated as discrete historical periods but as interpretive groupings derived from observable changes in publication volume and thematic emphasis.

- The period from 2014 to 2017 represents an early consolidation phase, during which research output remained relatively modest. Studies published in this phase primarily focused on formative assessment practices, including portfolio use, peer feedback, and reflective activities, often situated within classroom-based pedagogical reform.
- From 2018 to 2020, publication activity increased more rapidly, reflecting diversification in research focus. This phase is characterised by the emergence of studies examining hybrid assessment models, digital feedback tools, and teacher assessment literacy, suggesting growing professional and institutional engagement with alternative assessment practices.

c) The period from 2021 to 2024 shows a marked acceleration in publication volume. This increase coincides with the widespread digitalisation of higher education following the COVID-19 pandemic. Research during this phase demonstrates heightened attention to online formative assessment, e-portfolios, and technology-mediated feedback, including exploratory work on AI-supported assessment tools. While empirical output expanded substantially, this phase also reveals increasing variability in conceptual framing, with methodological innovation often outpacing theoretical integration.

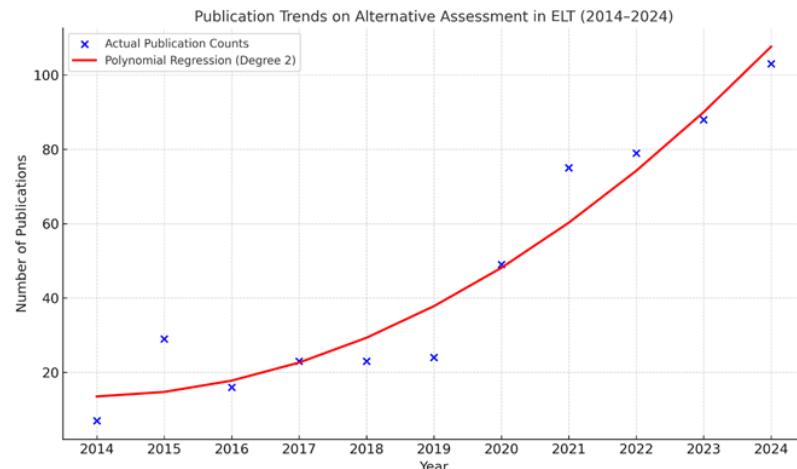


Figure 2 Publication trend of alternative assessment studies in ELT higher education (2014–2024)

2. Thematic Structure: Co-Word Clustering

The co-word analysis yielded five major thematic clusters representing recurrent areas of emphasis within alternative assessment research in ELT higher education. These clusters reflect patterns of term co-occurrence within titles and abstracts and are interpreted as indicative of shared conceptual space rather than as mutually exclusive categories. Figure 3 presents the co-word network visualisation, with clusters differentiated by colour.

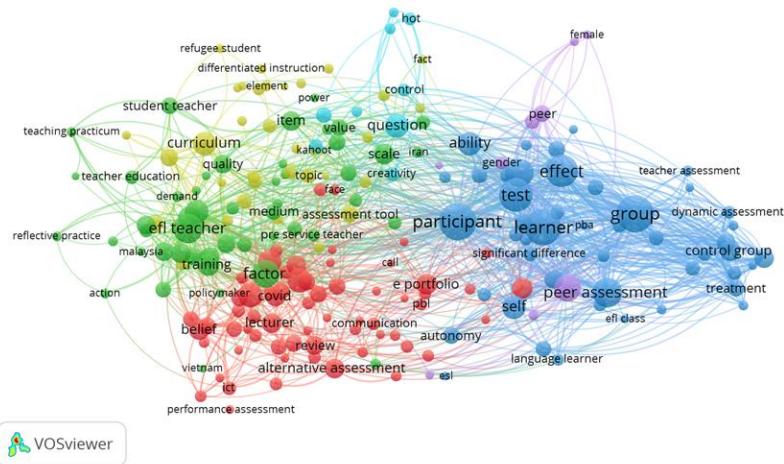


Figure 3 Co-word map of alternative assessment research in ELT higher education

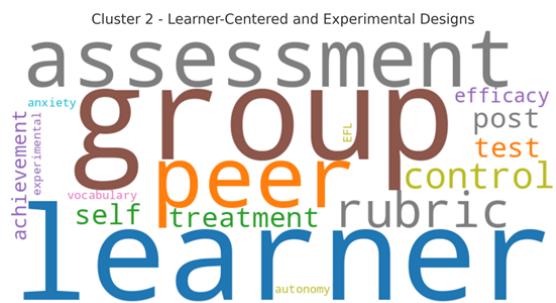
The co-word analysis identified five major thematic clusters representing dominant conceptual orientations within alternative assessment research in English Language Teaching (ELT) higher education. These clusters reflect patterns of term co-occurrence in article titles and abstracts and are interpreted as shared conceptual spaces rather than discrete or hierarchical categories.

a. Cluster 1: Assessment literacy and professional engagement



Cluster 1 is characterised by frequent co-occurrence of terms related to teacher assessment literacy, feedback practices, and professional learning. This pattern reflects sustained scholarly attention to educators' understanding of assessment principles and their capacity to implement formative and alternative assessment practices (Scarino, 2013; Wiliam, 2011). In university ELT contexts, a corresponding teaching application would include lecturers engaging in assessment literacy initiatives and redesigning writing assessments to incorporate portfolio-based tasks and iterative feedback cycles, supporting transparent criteria use and longitudinal monitoring of student development.

b. Cluster 2: Learner-Centred Experimental Designs



Cluster 2 shows a strong association with terms such as *peer assessment*, *rubrics*, *learning outcomes*, and *reliability*, indicating a focus on empirically evaluating alternative assessment practices using experimental or quasi-experimental designs (Panadero et al., 2016; Kane, 2013). A university-level ELT practice aligned with this cluster is structured peer review in academic writing or speaking courses, where students apply analytic rubrics to assess peers' drafts or presentations, generating outcome data that allow comparison across cohorts or instructional conditions.

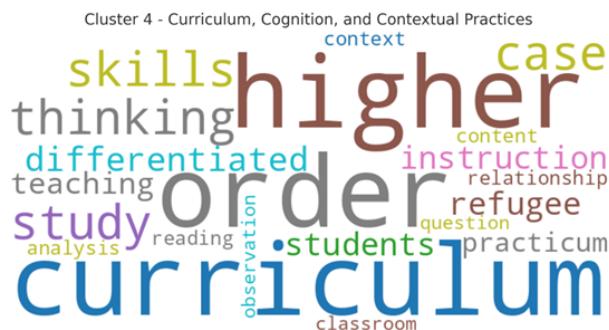
c. Cluster 3: Systemic Reform and Teacher Beliefs



This cluster is defined by co-occurring terms related to institutional policy, curriculum reform, and teacher beliefs, highlighting interest in how macro-level structures influence assessment practices (Shavelson & Towne, 2002; Cowie & Bell, 1999). In ELT programmes at the university level, an aligned example would include faculty-wide requirements to embed formative assessment components across courses, supported by programme guidelines, moderation procedures, or curriculum mapping

exercises. Co-word patterns suggest that such studies foreground implementation contexts more strongly than evidential justification processes.

d. Cluster 4: Curriculum and Inclusivity



Cluster 4 foregrounds ethical and equity-related concepts, with frequent co-occurrence of terms such as *fairness*, *inclusivity*, *diversity*, and *differentiated assessment*. This orientation aligns with scholarship emphasising fairness and social consequences as central dimensions of assessment quality (Shepard, 2000; Kunnan, 2017). In university ELT teaching, a practice corresponding to this cluster would involve offering multiple assessment formats, for example, written assignments, oral presentations, and collaborative projects, enabling students to demonstrate achievement through varied modes while aligning with shared learning outcomes (Shohamy, 2020). Co-word patterns indicate that inclusivity is discussed primarily at the level of design principles rather than through explicit inference-based validation language.

e. Cluster 5: Peer Identity and Sociocultural Feedback Dynamics



The fifth cluster is marked by the co-occurrence of terms associated with peer feedback, dialogue, collaboration, and learner identity, reflecting a sociocultural orientation toward assessment as a relational and interactional process (Vygotsky, 1978; Lantolf & Poehner, 2011). In university ELT classrooms, a corresponding practice would include peer-feedback workshops following oral presentations or draft submissions, where students negotiate criteria and meaning through guided dialogue (Carless & Boud, 2018). While conceptually cohesive, this cluster shows limited linkage to methodological or generalisation-oriented terms, consistent with critiques of sociocultural assessment research for limited engagement with formal validation procedures (Phongsirikul, 2018).

Across clusters, co-word distributions indicate uneven engagement with evidential, interpretive, and consequential dimensions of assessment. Measurement-oriented clusters demonstrate a stronger association with outcome and evaluation terms, whereas sociocultural clusters are more closely linked to interactional and identity-related constructs. Policy- and equity-oriented clusters emphasise consequences and access but show a weaker connection to inferential terminology. These bibliometric patterns provide empirical grounding for the subsequent discussion of relational validity

as an integrative framework capable of connecting evidence, interpretation, and educational consequence in alternative assessment research (Messick, 1989; Kane, 2013).

DISCUSSION

The findings of this bibliometric analysis indicate that research on alternative assessment in English Language Teaching (ELT) higher education has expanded rapidly over the past decade, yet has developed along largely parallel conceptual trajectories. The co-word patterns identified across clusters reveal uneven engagement with evidential, interpretive, and consequential dimensions of assessment quality. Measurement-oriented strands prioritise outcomes and evaluative consistency, sociocultural strands foreground interaction and identity, and policy- and equity-oriented strands emphasise access and consequence. While each orientation contributes important insights, their limited integration suggests that alternative assessment research remains conceptually segmented.

This segmentation reflects a longstanding epistemological divide in assessment scholarship. Validation-oriented traditions conceptualise assessment quality primarily through evidential reasoning, focusing on reliability, inference, and defensibility of decisions (Messick, 1989; Kane, 2013). In contrast, sociocultural perspectives position assessment as a mediated and dialogic practice, embedded in learning activity and shaped by relationships, identity, and participation (Vygotsky, 1978; Lantolf & Poehner, 2011; Scarino, 2013). The present findings suggest that these traditions continue to inform distinct strands of alternative assessment research in ELT higher education, with limited cross-fertilisation.

The cluster focusing on learner-centred experimental designs demonstrates strong engagement with evidential concerns, particularly through the use of rubrics, peer assessment protocols, and quasi-experimental designs. However, the relative absence of sociocultural constructs within this strand suggests that learning outcomes are often examined without sustained attention to how meaning, identity, and evaluative judgement are co-constructed in assessment processes. Conversely, clusters emphasising peer feedback and identity foreground relational and interpretive dimensions of assessment but show weaker linkage to inferential terminology and generalisation-oriented concepts. This imbalance constrains the extent to which rich, context-sensitive insights can be translated into defensible assessment interpretations beyond local settings.

Policy and inclusivity-oriented clusters highlight the importance of educational consequences, particularly fairness, access, and responsiveness to learner diversity. These studies foreground ethical considerations and institutional reform, aligning with broader calls to re-centre assessment within a learning culture rather than a purely measurement-driven paradigm (Shepard, 2000; Shohamy, 2020). However, the bibliometric patterns indicate that discussions of consequence are often decoupled from explicit evidential reasoning, resulting in fairness being articulated as a principle rather than as an analytically grounded component of validity.

Taken together, these patterns point to what can be characterised as a validity gap in alternative assessment research. Validity is frequently invoked but often operationalised narrowly, equated with reliability or procedural consistency, or treated implicitly within pedagogical or ethical discourse. This finding aligns with critiques that classroom-based and alternative assessments pose challenges for traditional validation frameworks, particularly when evidence is longitudinal, dialogic, and embedded in pedagogical relationships (Phakiti & Isaacs, 2021; Taylor & Banerjee, 2023). The absence of integrative frameworks capable of connecting evidence, interpretation, and consequence limits the field's capacity to articulate coherent accounts of assessment quality.

In response to this gap, the present study advances relational validity as a synthesising interpretive framework. Relational validity conceptualises validity as coherence among empirical evidence, contextual interpretation, and educational consequence. Building on argument-based validity theory (Messick, 1989; Kane, 2013) and sociocultural perspectives on assessment as mediated practice (Scarino, 2013; Shepard, 2019), this framework reframes assessment quality as an ongoing, situated process rather than a fixed procedural endpoint. Reliability is understood as contextual consistency supported through shared criteria and stable judgment practices, while fairness emerges through

participatory assessment cultures that enable learners to engage with standards and understand evaluative decisions (Kunnan, 2017).

The bibliometric findings provide empirical support for the need for such an integrative framework. The observed separation between evidential, interpretive, and consequential emphases across clusters suggests that no single strand sufficiently addresses assessment quality in its entirety. Relational validity offers a conceptual bridge by recognising that assessment interpretations must be justified not only through technical adequacy but also through their alignment with pedagogical intent and ethical consequence. This perspective is particularly salient in ELT higher education, where assessment practices increasingly involve peer interaction, iterative feedback, and diverse learner populations.

From a research perspective, these findings suggest the need for studies that explicitly articulate how alternative assessment practices generate evidence, how that evidence is interpreted within specific contexts, and how resulting decisions affect learners. Rather than positioning validation and sociocultural inquiry as competing paradigms, future research would benefit from designs that examine their interaction. For example, studies of peer assessment could integrate argument-based validation frameworks to explicate inferential warrants while retaining attention to dialogue, identity, and mediation. Similarly, inclusivity-oriented research could strengthen claims about fairness by systematically examining how assessment interpretations are supported and contested in practice.

In sum, the present study contributes to the field by empirically mapping the conceptual landscape of alternative assessment research in ELT higher education and by proposing relational validity as a theoretically grounded response to observed fragmentation. By situating validity at the intersection of evidence, interpretation, and educational consequence, relational validity provides a coherent lens for advancing assessment practices that are not only methodologically sound but also pedagogically meaningful and ethically responsive.

The findings of this study have several implications for research, teaching practice, and policy in English Language Teaching (ELT) higher education.

1. Implications for Research

For researchers, the findings underscore the importance of moving beyond parallel theoretical traditions toward more integrative research designs. Future studies on alternative assessment should explicitly articulate how assessment evidence is generated, how interpretations are warranted, and how decisions affect learners across contexts. Rather than treating validity, fairness, and learning impact as separate concerns, researchers are encouraged to examine their interaction within specific assessment practices.

The concept of relational validity provides a useful lens for this purpose. Researchers investigating peer assessment, portfolios, or formative feedback could combine sociocultural analysis of interaction and mediation with argument-based validation approaches that make inferential reasoning explicit. Such integration would strengthen the defensibility of claims while preserving sensitivity to context, identity, and learning processes. Longitudinal and mixed-method designs are particularly well suited to examining how evidential consistency, interpretive judgement, and educational consequences evolve.

In addition, the bibliometric patterns suggest a need for greater conceptual clarity in the use of assessment-related terminology. Researchers should define how constructs such as alternative, formative, authentic, and performance-based assessment are operationalised within their studies, thereby reducing ambiguity and enhancing cumulative knowledge building in the field.

2. Implications for ELT Teaching Practice

For ELT lecturers in higher education, the findings highlight the importance of developing assessment literacy that extends beyond technical design toward interpretive and ethical reasoning. Assessment decisions in alternative formats, such as peer assessment or portfolio-based evaluation, require lecturers to justify how evidence is interpreted and how feedback and grades influence learner development. Relational validity reframes assessment literacy as the capacity to align criteria, judgment, and consequence within specific teaching contexts.

In practice, this implies that alternative assessment tasks should be designed with explicit attention to transparency of criteria, opportunities for learner engagement with standards, and consistency of judgement across assessors and occasions. For example, peer-feedback activities should be accompanied by scaffolded use of rubrics and guided reflection to support shared understanding of quality. Similarly, portfolio assessment should incorporate clear rationales for evidence selection and decision-making processes, enabling students to understand how their work is evaluated over time.

In increasingly digital and blended learning environments, ELT lecturers must also balance the use of technological tools with human interpretive judgement. While digital platforms and AI-supported feedback systems can enhance efficiency and access, relational validity emphasises that interpretive responsibility remains with educators. Technology should support, rather than replace, dialogic feedback and contextualised judgement.

3. Implications for Policy and Institutional Practice

At the institutional level, the findings suggest that assessment reform initiatives should be evaluated not only in terms of compliance or standardisation but also in terms of coherence among evidence, interpretation, and educational consequence. Policies promoting alternative or formative assessment should be accompanied by support structures that develop teachers' assessment literacy and provide space for contextual adaptation.

In ELT programmes, particularly within Asian and Open and Distance Learning contexts, policymakers should recognise that fairness and inclusivity cannot be ensured solely through uniform procedures. Relational validity highlights the need for policies that allow flexibility in assessment formats while maintaining transparent criteria and shared standards. Moderation processes, professional learning communities, and programme-level dialogue about assessment interpretation can support this balance.

Furthermore, institutions should adopt mechanisms for ongoing evaluation of assessment practices. Rather than treating validity as a one-time technical check, relational validity frames it as an ongoing process requiring reflection on how assessment decisions shape learner trajectories and opportunities. Embedding such reflective practices within quality assurance systems can enhance both accountability and educational responsiveness.

CONCLUSION

The implications of this study point toward a shift in how assessment quality is conceptualised and enacted in ELT higher education. By adopting relational validity as an integrative framework, researchers, educators, and policymakers can move toward assessment practices that are evidentially defensible, pedagogically meaningful, and ethically responsive. This shift is particularly timely in contexts characterised by diverse learner populations, digital mediation, and expanding use of alternative assessment formats.

ACKNOWLEDGEMENT

The authors would like to thank the Wawasan Open University for providing the support and facilities.

FUNDING

The authors declare that no financial support was received for the research, authorship, and publication of this article.

DATA AVAILABILITY STATEMENT

Data will be made available on request.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the author(s) used ChatGPT to brainstorm ideas and beautify the paragraphs. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

Altinay, F., Alpturk, K., Berigel, M., Ayaz, A., Dagli, G., & Altinay, Z. (2024). Evolution of online assessment in higher education: A bibliometric analysis. *Pakistan Journal of Life and Social Sciences*, 22(1), 6772–6786. <https://doi.org/10.57239/PJLSS-2024-22.1.00494>

American Educational Research Association. (2011). Code of ethics. <https://www.aera.net/About-AERA/AERA-Rules-Policies/Professional-Ethics>

Andrade, H., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (Vol. 4, pp. 293–309). Springer. https://doi.org/10.1007/978-3-319-39211-0_17

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & evaluation in higher education*, 31(4), 399-413. <https://doi.org/10.1080/02602930600679050>

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–675. <https://www.jstor.org/stable/3587999>

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. Routledge. <https://doi.org/10.4324/9781315740621>

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>

Cheng, L., & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student learning*. Bloomsbury Publishing.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101-116. <https://doi.org/10.1080/0969594993026>

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. Routledge.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge.

Gipps, C. (1994). Developments in Educational Assessment: what makes a good test? *Assessment in Education: Principles, Policy & Practice*, 1(3), 283-292. <https://doi.org/10.1080/0969594940010304>

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>

Hook, D. W., Porter, S. J., & Herzog, C. (2018). Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, 3(23), 1-11. <https://doi.org/10.3389/frma.2018.00023>

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.

Kunnan, A. J. (2017). *Evaluating language assessments*. Routledge. <https://doi.org/10.4324/9780203803554>

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 11–33. <https://doi.org/10.1177/1362168810383328>

Lin, Y., & Yu, Z. (2023). A bibliometric analysis of peer assessment in online language courses. *Languages*, 8(1), 47. <https://doi.org/10.3390/languages8010047>

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351–372. <https://doi.org/10.1177/026553220101800403>

Messick, S. (1989). Validity. In R. Linn (Eds.), *Educational measurement* (pp. 13–103). Macmillan Publishing Co, Inc; American Council on Education.

Messick, S. (1996). Validity in performance assessments. In G. W. Phillips (Eds.), *Technical issues in large-scale performance assessment*. Washington DC: National Center for Education Statistics.

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation*, 13(1), 4. <https://doi.org/10.7275/0qpc-ws45>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71), 1-9. <https://doi.org/10.1136/bmj.n71>

Panadero, E., Brown, G. T., & Strijbos, JW. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803–830. <https://doi.org/10.1007/s10648-015-9350-2>

Phakiti, A., & Isaacs, T. (2021). Classroom assessment and validity: Psychometric and edumetric approaches. *European Journal of Applied Linguistics and TEFL*, 10(1), 3–24.

Phongsirikul, M. (2018). Traditional and alternative assessments in ELT: Students' and teachers' perceptions. *rEFLections*, 25(1), 61–84. <https://doi.org/10.61508/refl.v25i1.136267>

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed>

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 183–200. <https://doi.org/10.1177/0002716219843818>

Shohamy, E. (2020). *The power of tests: A critical perspective on the uses of language tests*. Routledge.

Taylor, L., & Banerjee, J. (2023). Language assessment accommodations: Issues and challenges for the future. *Language Testing*, 40(4), 1000–1006. <https://doi.org/10.1177/02655322231186222>

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523–538. <https://doi.org/10.1007/s11192-009-0146-3>

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. https://doi.org/10.1162/qss_a_00112

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, (Eds.). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>

Zhang, X. (2025). Trends in language assessment and testing: A bibliometric study. *Studies in Second Language Learning and Teaching*, 15(1), 171–198. <https://doi.org/10.14746/ssllt.25141>