European Journal of Applied Linguistics Studies
ISSN: 2602 - 0254
ISSN-L: 2602 - 0254
Available on-line at: http://www.oapub.org/lit

DOI: 10.46827/ejals.v9i1.698 Volume 9 | Issue 1 | 2026

INTERLANGUAGE DEVELOPMENT IN L2 ACADEMIC WRITING: A STATE-OF-THE-ART REVIEW OF MEASUREMENT, TASK DESIGN, AND WRITTEN FEEDBACK IN EMI HIGHER EDUCATION

Phuong Bao Tran Nguyen[i]
School of Foreign Language,
Can Tho University,
Vietnam

Abstract:
English-medium instruction (EMI) has expanded rapidly in higher education; however, programs often lack defensible ways to monitor students' second-language (L2) academic writing development beyond course grades or ad hoc teacher judgments. This review synthesises key theoretical and methodological debates shaping how interlanguage development in academic writing is conceptualised and assessed in EMI settings. We integrate (i) interlanguage and writing-process perspectives on attention, working memory, and developmental constraints; (ii) measurement traditions centred on complexity, accuracy, and fluency (CAF) and their extension to functional adequacy (FA) to capture task- and genre-appropriate communicative success; (iii) the increasing use of automated indices (e.g., lexical sophistication and cohesion tools) and the validity risks they pose when genre control and interpretive safeguards are weak; and (iv) task design and written corrective feedback (WCF) as mechanisms that can either exacerbate or mitigate the "EMI plateau," where content load crowds out language attention. Across these strands, we argue that EMI writing development is best understood as an interaction among task demands, attentional allocation, and feedback-mediated revision opportunities, and that robust program monitoring requires construct coverage (CAF plus FA), genre-sensitive task comparability, and transparent interpretive rules for automated metrics. We conclude by outlining priority controversies, research gaps, and near-term developments, including minimum reporting standards for EMI writing studies and scalable, low-stakes assessment architectures that better align language development with disciplinary writing demands.

Keywords: English-medium instruction; interlanguage; L2 academic writing; CAF; functional adequacy; task complexity; written corrective feedback; automated writing assessment

[i] Correspondence: email npbtran@ctu.edu.vn

1. Introduction

EMI has become a defining feature of higher education internationalization, including in contexts where students and lecturers operate largely outside English-dominant environments (Mckinley et al., 2022; Tai & Zhao, 2022; Vidal & Jarvis, 2018). While EMI aims to broaden access to disciplinary knowledge and global academic participation, it also raises persistent questions about whether and how students' academic English develops through content study, particularly for their writing skills (Aizawa et al., 2025; Kamaşak & Sahan, 2023; Nguyen & Sercu, 2021; Zhou et al., 2023).

How EMI programs can measure and judge whether students' academic writing is actually developing over time. Many EMI programs lack a coherent measurement framework for tracking L2 academic writing development over time, and longitudinal evidence remains comparatively limited in the broader EMI evidence base (Curle et al., 2024; Liao et al., 2025; Rose et al., 2021; Xu et al., 2021). In practice, academic writing is often assessed through course grades, isolated assignments, or teacher feedback that is not comparable across semesters, genres, or disciplines—limiting inference about development rather than task-specific performance (Galloway & Rose, 2021). This makes it difficult to distinguish genuine interlanguage development from task- and genre-contingent performance and to diagnose which instructional design features sustain or constrain progress (Adams & Guillot, 2008; Manchón & Sanz, 2023; Mccutchen, 1996).

This review addresses this gap by consolidating major schools of thought and methodological choices that shape claims about writing development in EMI higher education. We focus on (i) how interlanguage development is conceptualized in relation to writing processes (attention, working memory, planning, and revision) , (ii) how development is operationalized through linguistic and functional measures, and the known tensions in CAF-only operationalizations, (iii) what automated indices can and cannot validly contribute to monitoring, and (iv) how task design and written corrective feedback interact with cognitive load to influence development trajectories .

As a review article, this paper provides an integrative synthesis of seminal and representative scholarship across second language acquisition, L2 writing, and EMI research. Rather than enumerating an exhaustive search and screening workflow, we organize evidence around recurring controversies, convergent findings, and design implications for EMI writing research and program practice. We pay particular attention to contexts similar to those of Vietnamese public universities, where resources and assessment infrastructure are often constrained (Nguyen, 2018; Nguyen et al., 2016; Tran & Nguyen, 2018).

We advance three claims. First, EMI writing development should be interpreted as a dynamic allocation problem in which task demands, disciplinary content load, and feedback opportunities shape the attention available for language learning (Richards & Pun, 2021; Zhou et al., 2023). Second, measurement should move beyond CAF-only batteries to include functional adequacy (FA) when making claims about development in academic genres (Fang, 2021; Kuiken & Vedder, 2018; Rakedzon & Baram-Tsabari, 2016).

Third, automated indices can support monitoring only when embedded in genre-controlled designs and interpreted with explicit validity safeguards (Barasa, 2024; Bouwer et al., 2014).

2. Conceptual foundations

Interlanguage development in writing reflects changes in learners' linguistic resources and their ability to deploy those resources under task constraints, often evidenced through multidimensional performance profiles (complexity, accuracy, and fluency) rather than uniform gains across all dimensions (Abrams & Rott, 2016; Götz et al., 2022; Luo & Sun, 2025; Rahimi & Zhang, 2018). In writing, development is inseparable from process: planning, formulation or translation, monitoring, and revision compete for limited attentional capacity and working memory resources (Al Asmari, 2013; Hayes & Chenoweth, 2006; T Kellogg et al., 2013). Writing models that foreground working memory and executive control help explain why learners may show uneven progress across dimensions (e.g., gains in lexical or syntactic sophistication without parallel gains in accuracy) and why performance can regress under higher cognitive or conceptual load (Al-Shehri & Gitsaki, 2010).

Within second language acquisition (SLA) and task-based perspectives, two influential positions frame expectations about development under increasing task demands (Jin & Yan, 2024). Trade-off perspectives predict that limited attentional resources force learners to prioritise certain dimensions (e.g., fluency or complexity) at the expense of others (often accuracy), producing systematic CAF trade-offs under time pressure or increased conceptual demands (Jin & Yan, 2024). In contrast, cognition-oriented accounts argue that appropriately sequenced increases in task complexity can stimulate deeper processing and promote development—especially in complexity and accuracy—provided that learners have sufficient support (e.g., planning time, scaffolding, and feedback) to allocate attention to form as well as meaning (Galloway & Rose, 2021; Richards & Pun, 2021; Zhou et al., 2023).

In EMI contexts, these debates take on added salience because disciplinary learning is not merely "topic content"; it is conceptually dense, assessment-driven, and often delivered at pace, with recurring tensions around what is assessed (content vs language) and who is responsible for supporting language development (Dimova & Kling, 2018; Zhou et al., 2023) . Large-scale and multi-context EMI research also indicates that writing and speaking frequently emerge as "pinch points" of difficulty, with variation by discipline and learner background (Aizawa et al., 2025; Kamaşak & Sahan, 2023; S. Zhou et al., 2023). If EMI tasks systematically overload learners' resources, a plateau can emerge in which students maintain intelligible writing but stabilize below discipline-appropriate accuracy, rhetorical control, and register—an outcome that is plausibly attributable to the interaction between task demands, limited support, and feedback ecology rather than to learner deficit alone (Abdi Tabari & Huang, 2025; Brudermann et al., 2021; Lee, 2018).

3. Measuring interlanguage development in EMI writing

Most empirical work operationalizes writing development through complexity, accuracy, and fluency (CAF) (Barrot & Gabinete, 2019; Plakans et al., 2016). CAF measures provide a tractable way to quantify change, but they can also encourage narrow construct coverage when applied without genre sensitivity and an explicit evidentiary logic (Biber et al., 2011; Imaz Agirre et al., 2024). For example, clause-based syntactic complexity indices may under-represent development in academic writing where progress is often expressed through phrasal elaboration (e.g., complex noun phrases), nominal styles, and informational density rather than increased clausal subordination (Deng et al., 2020; Kyle & Crossley, 2018; Lambert & Nakamura, 2018).

An important extension is the inclusion of functional adequacy (FA), which captures how well a text accomplishes communicative goals in relation to the task and genre (e.g., content fulfillment, task requirements, comprehensibility, coherence, cohesion) (Kuiken & Vedder, 2018, 2019, 2022). In EMI settings, FA is not optional: the point of writing is to demonstrate disciplinary meaning-making, and students may produce linguistically "advanced" texts that remain functionally misaligned with academic genre expectations—underscoring the need to keep "adequacy" conceptually distinct from CAF while using it to interpret CAF patterns (Dafouz et al., 2018; Kuiken & Vedder, 2019).

A combined measurement logic (CAF and FA) supports more defensible developmental interpretations: CAF can index changes in linguistic resources, while FA situates those resources within task success and genre-appropriate communication (Kuiken & Vedder, 2022; Pallotti, 2009). However, the combined approach requires methodological discipline: stable or comparable tasks, explicit rating rubrics, rater training, and transparent reporting of inter-rater reliability (Wind & Peterson, 2017), particularly because FA judgments are inherently interpretive and may involve both subject lecturers and language instructors.

Finally, developmental claims should distinguish between product change (text-level outcomes) and process change (planning, revision behavior, uptake of feedback). Product-only outcomes can mask meaningful shifts in writers' online allocation of attention and revision strategies (Hedgcock & Lefkowitz, 1996; Newman et al., 2013; Sugita, 2006), and classic revision frameworks emphasize that revisions differ in depth—ranging from surface edits to meaning-changing (text-based) revisions (Chang, 2025). EMI programs that seek sustainable improvement may therefore benefit from integrating process indicators—such as revision depth, pausing or revision patterns, or feedback uptake—alongside text metrics, while acknowledging feasibility limits of large-scale process data collection (Agarwal, 2024; Ahmed et al., 2024; Nguyen et al., 2017; Thi & Nikolov, 2021).

4. Automation and extended indices: promise and validity constraints

Automated indices and tools for lexical sophistication, syntactic complexity, cohesion, and readability have expanded the feasible measurement space for EMI programs and researchers(Kyle & Crossley, 2014; Lu, 2011; Lu & Hu, 2021; Mcnamara et al., 2009; Vidal & Jarvis, 2018). These tools can reduce scoring burden, enable repeated measurement, and support diagnostic feedback at scale. Nonetheless, their use in EMI writing research is constrained by validity risks that are often under-discussed in program-facing applications. First, automated scores are sensitive to genre, prompt, and text length; without careful task comparability, differences in metrics may reflect topic or genre shifts rather than development (Bouwer et al., 2014). Second, some indices have ambiguous interpretability in advanced academic writing. For example, explicit cohesive devices or lexical overlap may decrease as writers adopt more implicit and discipline-typical cohesion strategies; validation work on cohesion tools shows that some local cohesion indices are weakly related—or even negatively related—to expert judgments of coherence/quality, depending on the cohesion dimension and writing context (Castro, 2004; Connor, 1984; Johnson, 1992). Third, tool outputs can be overinterpreted as proxies for "quality" unless triangulated with human judgment and anchored to genre expectations—an issue repeatedly emphasized in the automated writing evaluation literature (Crossley & Mcnamara, 2011; Goldshtein et al., 2024; Zedelius et al., 2018).

Accordingly, automation should be embedded within an explicit validity argument: specify the constructs targeted, justify why selected indices represent them for the given genre, control for confounds (length, prompt, register), and predefine interpretive rules for what constitutes meaningful change (Chapelle et al., 2015; Kane, 2004). For program monitoring, a small, stable set of validated indices may be preferable to expansive batteries that encourage exploratory interpretation (Chen et al., 2016; Mohsen, 2022).

5. Task design and written corrective feedback as developmental mechanisms

Task design in EMI can shape development by modulating cognitive load and the opportunities learners have to attend to language form (Coleman et al., 2018; Xie & Curle, 2020). In EMI, tasks frequently prioritize content mastery and assessment demands; writing is used as an instrument for demonstrating knowledge, and instructors often report tensions around the extent to which language development can be addressed within content-driven courses (Avargil et al., 2011; Massler et al., 2014; Saricaoglu & Bilki, 2021). This raises the risk that learners allocate attention to idea generation and disciplinary correctness while leaving limited capacity for grammatical accuracy, lexical precision, and rhetorical refinement—a pattern consistent with limited-capacity accounts of performance under meaning-focused demands (Graham, 2016; Macaro et al., 2017; Rahman & Mehar Singh, 2019).

Task complexity research suggests that developmental benefits depend on how complexity is manipulated and what support is provided. When conceptual complexity rises without scaffolding, learners may sacrifice accuracy and control, consistent with predicted trade-offs among complexity, accuracy, and fluency under attentional constraints (Bui & Skehan, 2018; Jin & Yan, 2024; Khezrlou, 2020; Robinson, 2007). When increases are sequenced and accompanied by planning, models, and targeted feedback, learners may engage in deeper processing that supports durable gains—an expectation central to cognition-oriented accounts of task sequencing (Cerezo et al., 2019; Khezrlou, 2020; Tan et al., 2022).

Written corrective feedback (WCF) is a key lever for redirecting attention to form and genre-appropriate discourse (Abdi Tabari & Huang, 2025; Brudermann et al., 2021; J. Lee, 2018). Evidence across L2 writing indicates that WCF effects are conditional: they depend on feedback focus (focused vs. unfocused), opportunities for revision, learner readiness, and the alignment between feedback targets and task demands(Chen et al., 2016; Karim & Nassaji, 2018; Tsao, 2021). Debate in the field also underscores that effectiveness claims hinge on design features such as sustained treatment, revision cycles, and appropriate outcome measures (Carless et al., 2020; Sykes et al., 2024). In EMI, feedback also functions as mediation between language and disciplinary expectations; therefore, coordination between language teachers and subject lecturers is likely to be consequential for functional adequacy outcomes and for the consistency of expectations communicated to students (Cameron et al., 1996; Wingate et al., 2011).

For EMI programs, a pragmatic implication is that development is more plausible when tasks are designed with "attention windows": structured planning, manageable conceptual scope, and iterative drafting cycles in which WCF targets a stable set of high-leverage features (Lasagabaster, 2022; Lin & Lei, 2021; Macaro, 2022). This design logic reframes the plateau as an instructional ecology problem—arising from recurring task demands, support conditions, and feedback practices—rather than a fixed limit in learner capacity (Mckinley, 2022; Sahan et al., 2022; Tran & Nguyen, 2018).

6. EMI higher education practice: implications for contexts such as Vietnam

EMI in many non-Anglophone universities operates under resource and infrastructure constraints—large cohorts, limited contact time for sustained academic English support, and assessment systems primarily oriented toward disciplinary content. In Vietnam, empirical accounts from EMI lecturers and students repeatedly highlight constraints related to English proficiency gaps, pedagogical and assessment pressures, and resource availability, all of which can make writing development uneven and weakly monitored (H. T. Nguyen et al., 2016; H. T. Nguyen, 2018; Tran & Nguyen, 2018). For Vietnamese public universities, practical solutions therefore need to be scalable, low-stakes, and curriculum-compatible, rather than dependent on intensive one-to-one writing support or high-cost testing infrastructures (Sahan et al., 2022).

Two implementation principles follow from the synthesis above. First, monitoring should align with disciplinary genre demands. A pragmatic approach is to select a small set of genre families (e.g., explanation, report, argument) that recur across the program and anchor both measurement and feedback to stable genre expectations; genre-family classifications of assessed university writing provide a defensible basis for this kind of mapping across disciplines (Ali et al., 2025; Fang Xu, 2024; Otokiti et al., 2022). Second, assessment should prioritize interpretability and actionability. Rather than maximizing score precision, feedback loops should make standards visible and specify "what to do next" (e.g., strengthen claim–evidence links, increase phrasal elaboration, tighten stance marking), consistent with formative assessment principles emphasizing usable criteria, calibrated standards, and feedback that supports self-regulation (Kibble, 2017; Kustitskaya et al., 2022; Williamson et al., 2004) .

While comprehensive longitudinal tracking may be aspirational, EMI programs can adopt stepwise monitoring architectures: periodic benchmark tasks using stable rubrics, sampled portfolios, or rotating cohort assessments that distribute workload while preserving comparability over time. Portfolio assessment scholarship in writing highlights both feasibility advantages (sampling across tasks or genres) and the need for explicit scoring procedures to manage reliability and interpretability (Brindley, 2001; Gulati et al., 2014; Krebs et al., 2022; Lertsakulbunlue & Kantiwong, 2024; Mendonça et al., 2025). Importantly, any use of automated indices should be framed as diagnostic support rather than high-stakes evaluation, and interpreted in conjunction with human judgments of functional adequacy and genre alignment; the automated writing evaluation literature stresses that responsible use requires an explicit validity argument about interpretation and use, plus safeguards against construct under-representation and confounds (Bradford, 2016; Curle et al., 2024; Hultgren et al., 2022; Mckinley & Rose, 2022; Toh, 2019).

7. Discussion: controversies, gaps, and future developments

Three debates repeatedly shape interpretations of EMI writing development. The first concerns task complexity: whether increasing cognitive and conceptual demands primarily induces trade-offs (e.g., reduced accuracy and control under limited attentional capacity) or instead stimulates deeper processing that supports development. Accumulated task-based work indicates that both patterns are attested, and the decisive factor is often task design and support—especially sequencing, planning and scaffolding, and revision opportunities that protect attention to form alongside meaning (Richards & Pun, 2021b; S. Zhou et al., 2023). In EMI specifically, these tensions are amplified because conceptual density and assessment pressure can systematically redirect attention toward disciplinary content at the expense of language form, a recurring concern in the broader EMI evidence base (Curle, Alqarni, et al., 2024; Liao et al., 2025; Rose et al., 2021; Xu et al., 2021).

The second debate concerns measurement: whether CAF-only batteries adequately represent development in academic writing. While CAF remains a practical measurement tradition, scholars have long cautioned that CAF indices can drift from the construct of "development" unless operational definitions, tasks, and interpretations are made explicit (Norris & Ortega, 2009). Moreover, clause-based complexity metrics may under-represent development in academic prose where increased "compression" is realized through phrasal elaboration (e.g., complex noun phrases) rather than clausal subordination (Biber et al., 2020; Biber & Gray, 2015; Staples et al., 2016). This evidentiary logic motivates incorporating functional adequacy—how well a text accomplishes task or genre goals—using explicit rubrics and documented reliability, particularly in EMI where writing is evaluated as disciplinary meaning-making (Biber et al., 2011; Goldshtein et al., 2024; Peng & Xie, 2021).

The third debate concerns automation: whether tool-based indices provide valid evidence of development. Automated analyses can support scalable monitoring, but their interpretability is constrained by sensitivity to prompt/genre and text length, and by the risk that users treat outputs as proxies for "quality" absent triangulation (Phyak, 2023; Richards & Pun, 2021b; Tajik et al., 2022). Accordingly, automation is best embedded in an explicit validity argument that specifies constructs, controls confounds, and defines what counts as meaningful change for the target genre—an approach aligned with argument-based validation traditions and AWE-focused validity work (Chapelle et al., 2015; Ranalli et al., 2016; Rönnebeck et al., 2016).

Research gaps remain substantial. The literature is still limited by weak task or genre comparability across time, under-reporting of rater procedures for functional measures, and inconsistent construct definitions (Deng et al., 2020; Lewis, 1992; Sun & Révész, 2021; Trace et al., 2016). Many studies retain clause-heavy complexity indices that may miss phrasal and discourse-level development typical of advanced academic writing (Biber et al., 2011). Automated metrics are frequently used without explicit validity arguments or confound controls (e.g., length and prompt effects), limiting developmental inference (Biber et al., 2011, 2020). Finally, there remains a shortage of studies that connect linguistic indicators to instructional mechanisms—for example, how feedback is processed and retained, and how revision depth relates to improvement—despite foundational work showing that learner engagement and revision operations matter for outcomes (I. Lee, 2019; D. Li & Zhang, 2021; Ranalli, 2018).

A feasible development agenda is to converge on minimum reporting and design standards. At a minimum, EMI writing studies should report genre or task specifications, comparability decisions, timing and exposure, and scoring procedures (including reliability when human judgments are used) (Kuiken & Vedder, 2018, 2019). Measurement batteries should be parsimonious but construct-complete, combining CAF indicators with functional adequacy anchored to genre criteria, and interpreted with the definitional discipline urged in CAF methodology work (Kuiken & Vedder, 2022). For automation, the field would benefit from shared benchmark datasets and cross-

disciplinary validation studies using transparent validity arguments and confound controls, rather than expanding exploratory index batteries (Dai et al., 2024).

In summary, this synthesis prioritises conceptual integration and design implications over exhaustive coverage. Publication bias, uneven disciplinary representation, and variability in local EMI models constrain generalisation —patterns already noted in EMI review work (Tai, 2023; Wang et al., 2025). Nonetheless, the converging evidence supports a defensible claim: EMI programs can more plausibly support interlanguage development in academic writing when they treat development as an interaction among task demands, attentional allocation, and feedback-mediated revision opportunities, and when they align constructs, tasks, and feedback around genre-sensitive frameworks that combine CAF with functional adequacy while deploying automation only within explicit validity safeguards.

Acknowledgements

The author gratefully acknowledges the support of Can Tho University, Vietnam, for providing the academic environment that enabled the development of this review. The author also wishes to express sincere appreciation to her colleague, Dr. Tran Thi Thanh Quyen, for her thoughtful feedback during the final review process. Appreciation is likewise extended to the students whose insights and engagement have indirectly contributed to the ideas discussed in this paper.

Conflict of Interest Statement
The author declares no conflicts of interest.

About the Author
Phuong Bao Tran Nguyen is a lecturer at the School of Foreign Languages, Can Tho University, Vietnam. Her research focuses on English-medium instruction (EMI), second language writing, interlanguage development, and AI-enhanced language education. She is especially interested in academic writing development in higher education and the use of technology to support English language teaching and learning.

References

Abdi Tabari, M., & Huang, Z. (2025). Examining task type effects on L2 writing complexity: understanding the moderating roles of goal orientation and flow. *Innovation in Language Learning and Teaching*, *ahead-of-print* (ahead-of-print), 1–22. https://doi.org/10.1080/17501229.2025.2560103

Abrams, Z., & Rott, S. (2016). Variability and variation of L2 grammar: A cross-sectional analysis of German learners' performance on two tasks. *Language Teaching Research*, *21*(2), 144–165. https://doi.org/10.1177/1362168815627833

Adams, A.-M., & Guillot, K. (2008). Working Memory and Writing in Bilingual Students. *ITL - International Journal of Applied Linguistics*, *156*(156), 13–28. https://doi.org/10.2143/itl.156.0.2034417

Agarwal, G. (2024). Test Case Automation: Transforming Software Testing in the Digital Era. *International Journal of Computing and Engineering*, *6*(5), 52–58. https://doi.org/10.47941/ijce.2314

Ahmed, Z., Shanto, S. S., Rime, M. H. K., Morol, M. K., Fahad, N., Hossen, M. J., & Abdullah-Al-Jubair, M. (2024). The Generative AI Landscape in Education: Mapping the Terrain of Opportunities, Challenges, and Student Perception. *IEEE Access*, *12*, 147023–147050. https://doi.org/10.1109/access.2024.3461874

Aizawa, I., Rose, H., Thompson, G., & Mckinley, J. (2025). Content knowledge attainment in English medium instruction: Does academic English literacy matter? *Language Teaching Research*. https://doi.org/10.1177/13621688241304051

Al Asmari, A. (2013). Practices and Prospects of Learner Autonomy: Teachers' Perceptions. *English Language Teaching*, *6*(3). https://doi.org/10.5539/elt.v6n3p1

Ali, M., Harbieh, I., & Haider, K. H. (2025). Bytes versus brains: A comparative study of AI-generated feedback and human tutor feedback in medical education. *Medical Teacher*, *48*(1), 131–141. https://doi.org/10.1080/0142159x.2025.2519639

Al-Shehri, S., & Gitsaki, C. (2010). Online reading: a preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. *ReCALL*, *22*(3), 356–375. https://doi.org/10.1017/s0958344010000212

Avargil, S., Herscovitz, O., & Dori, Y. J. (2011). Teaching Thinking Skills in Context-Based Learning: Teachers' Challenges and Assessment Knowledge. *Journal of Science Education and Technology*, *21*(2), 207–225. https://doi.org/10.1007/s10956-011-9302-7

Barasa, D. (2024). Demystifying the Discourse: Techniques to Effective Academic Writing. *Journal of Research and Academic Writing*, *1*(1), 13–21. https://doi.org/10.58721/jraw.v1i1.571

Barrot, J., & Gabinete, M. K. (2019). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. *International Review of Applied Linguistics in Language Teaching*, *59*(2), 209–232. https://doi.org/10.1515/iral-2017-0012

Biber, D., & Gray, B. (2015). *Grammatical Complexity in Academic English*. Cambridge University. https://doi.org/10.1017/cbo9780511920776

Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Reppen, R., Staples, S., & Egbert, J. (2020). Exploring the longitudinal development of grammatical complexity in the disciplinary writing of L2-English university students. *International Journal of Learner Corpus Research*, *6*(1), 38–71. https://doi.org/10.1075/ijlcr.18007.bib

Bouwer, R., Béguin, A., Sanders, T., & Van Den Bergh, H. (2014). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Bradford, A. (2016). Toward a Typology of Implementation Challenges Facing English-Medium Instruction in Higher Education. *Journal of Studies in International Education*, *20*(4), 339–356. https://doi.org/10.1177/1028315316647165

Brindley, G. (2001). Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing*, *18*(4), 393–407. https://doi.org/10.1177/026553220101800405

Brudermann, C., Grosbois, M., & Sarré, C. (2021). Accuracy development in L2 writing: Exploring the potential of computer-assisted unfocused indirect corrective feedback in an online EFL course. *ReCALL*, *33*(3), 248–264. https://doi.org/10.1017/s095834402100015x

Bui, G., & Skehan, P. (2018). *Complexity, Accuracy, and Fluency*. 1–7. https://doi.org/10.1002/9781118784235.eelt0046

Cameron, L., Moon, J., & Bygate, M. (1996). Language Development of Bilingual Pupils in the Mainstream : How do Pupils and Teachers Use Language? *Language and Education*, *10*(4), 221–236. https://doi.org/10.1080/09500789608666710

Carless, D., To, J., Kwan, C., & Kwok, J. (2020). Disciplinary perspectives on feedback processes: towards signature feedback practices. Teaching in Higher Education, 28(6), 1158–1172. https://doi.org/10.1080/13562517.2020.1863355

Castro, C. D. (2004). Cohesion and the social construction of meaning in the essays of Filipino college students writing in l2 English. *Asia Pacific Education Review*, *5*(2), 215–225. https://doi.org/10.1007/bf03024959

Cerezo, L., Manchón, R. M., & Nicolás-Conesa, F. (2019). *What Do Learners Notice While Processing Written Corrective Feedback?* (pp. 171–185). Routledge. https://doi.org/10.4324/9781315165080-12

Chang, H. (2025). Investigating the Relation Between Peer Critiquing and Model Revisions in Modeling-Based Learning. *Science Education*. https://doi.org/10.1002/sce.70026

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385–405. https://doi.org/10.1177/0265532214565386

Chen, S., Nassaji, H., & Liu, Q. (2016). EFL learners' perceptions and preferences of written corrective feedback: a case study of university students from Mainland

China. *Asian-Pacific Journal of Second and Foreign Language Education*, *1*(1). https://doi.org/10.1186/s40862-016-0010-y

Coleman, J., Hultgren, K., Li, W., Tsui, C. C., & Shaw, P. (2018). Forum on English-medium Instruction. *TESOL Quarterly*, *52*(3), 701–720. https://doi.org/10.1002/tesq.469

Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Paper in Linguistics*, *17*(3), 301–316. https://doi.org/10.1080/08351818409389208

Crossley, S. A., & Mcnamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, *21*(2/3), 170. https://doi.org/10.1504/ijceell.2011.040197

Curle, S., Alqarni, O. M., Mahdi, H. S., Al-Nofaie, H. A., & Ali, J. K. M. (2024). Exploring the nexus of variables in English-medium instruction research: a comprehensive global meta-analysis. *International Journal of Multilingualism*, *22*(3), 1202–1227. https://doi.org/10.1080/14790718.2024.2377199

Curle, S., Rose, H., & Yuksel, D. (2024). English medium instruction in emerging contexts: An editorial introduction to the special issue. *System*, *122*, 103262. https://doi.org/10.1016/j.system.2024.103262

Dafouz, E., Hüttner, J., & Smit, U. (2018). New Contexts, New Challenges for TESOL: Understanding Disciplinary Reasoning in Oral Interactions in English-Medium Instruction. *TESOL Quarterly*, *52*(3), 540–563. https://doi.org/10.1002/tesq.459

Dai, D. W., Vu, T., Knoch, U., Lim, A. S., Malone, D. T., & Mak, V. (2024). Expanding Kane's argument-based validity framework: What can validation practices in language assessment offer health professions education? *Medical Education*, *58*(12), 1462–1468. https://doi.org/10.1111/medu.15452

Deng, Y., Lei, L., & Liu, D. (2020). Calling for More Consistency, Refinement, and Critical Consideration in the Use of Syntactic Complexity Measures for Writing. *Applied Linguistics*, *42*(5), 1021–1028. https://doi.org/10.1093/applin/amz069

Dimova, S., & Kling, J. (2018). Assessing English-Medium Instruction Lecturer Language Proficiency Across Disciplines. *TESOL Quarterly*, *52*(3), 634–656. https://doi.org/10.1002/tesq.454

Fang Xu, D. C. (2024). Design and Implementation of Machine Learning Algorithms in Automatic Grading of Students' Assignments. *Journal of Electrical Systems*, *20*(3s), 899–919. https://doi.org/10.52783/jes.1388

Fang, Z. (2021). *Demystifying Academic Writing*. Routledge. https://doi.org/10.4324/9781003131618

Galloway, N., & Rose, H. (2021). English medium instruction and the English language practitioner. *ELT Journal*, *75*(1), 33–41. https://doi.org/10.1093/elt/ccaa063

Goldshtein, M., Alhashim, A. G., & Roscoe, R. D. (2024). An NLP-Based Exploration of Variance in Student Writing and Syntax: Implications for Automated Writing Evaluation. *Computers*, *13*(7), 160. https://doi.org/10.3390/computers13070160

Götz, S., Wolk, C., & Jäschke, K. (2022). *Contextualizing fluency in advanced spoken learner language* (pp. 273–298). John Benjamins Company. https://doi.org/10.1075/scl.104.11got

Graham, S. (2016). Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching*, *50*(1), 107–119. https://doi.org/10.1017/s0261444816000306

Gulati, M., Ram, S. S., & Singh, A. (2014). *An in depth study into using EMI signatures for appliance identification*. *10*, 70–79. https://doi.org/10.1145/2674061.2674070

Hayes, J. R., & Chenoweth, N. A. (2006). Is Working Memory Involved in the Transcribing and Editing of Texts? *Written Communication*, *23*(2), 135–149. https://doi.org/10.1177/0741088306286283

Hedgcock, J., & Lefkowitz, N. (1996). Some Input on Input: Two Analyses of Student Response to Expert Feedback in L2 Writing. *The Modern Language Journal*, *80*(3), 287–308. https://doi.org/10.1111/j.1540-4781.1996.tb01612.x

Huang, Y.-P. (2018). Learner Resistance to English-medium instruction practices: a qualitative case study. *Teaching in Higher Education*, *23*(4), 435–449. https://doi.org/10.1080/13562517.2017.1421629

Hultgren, A. K., Owen, N., Shrestha, P., Kuteeva, M., & Mežek, Š. (2022). Assessment and English as a medium of instruction. *Journal of English-Medium Instruction*, *1*(1), 105–123. https://doi.org/10.1075/jemi.21019.hul

Imaz Agirre, A., Arias-Hermoso, R., & Ipiña, N. (2024). The effect of an intervention focused on academic language on CAF measures in the multilingual writing of secondary students. *International Review of Applied Linguistics in Language Teaching*, *63*(2), 1373–1396. https://doi.org/10.1515/iral-2023-0137

Jiang, A. L., & Zhang, L. J. (2019). Chinese students' perceptions of English learning affordances and their agency in an English-medium instruction classroom context. *Language and Education*, *33*(4), 322–339. https://doi.org/10.1080/09500782.2019.1578789

Jin, C., & Yan, J. (2024). The effects of task complexity and task sequencing on L2 performance: a systematic review. *The Language Learning Journal*, *53*(1), 114–141. https://doi.org/10.1080/09571736.2024.2345911

Johnson, P. (1992). Cohesion and Coherence in Compositions in Malay and English. *RELC Journal*, *23*(2), 1–17. https://doi.org/10.1177/003368829202300201

Kamaşak, R., & Sahan, K. (2023). Academic Success in English Medium Courses: Exploring Student Challenges, Opinions, Language Proficiency and L2 Use. *RELC Journal*, *55*(3), 705–720. https://doi.org/10.1177/00336882231167611

Kane, M. (2004). Certification Testing as an Illustration of Argument-Based Validation. *Measurement: Interdisciplinary Research and Perspectives*, *2*(3), 135–170. https://doi.org/10.1207/s15366359mea0203_1

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, *24*(4), 519–539. https://doi.org/10.1177/1362168818802469

Khezrlou, S. (2020). The role of task repetition with direct written corrective feedback in L2 writing complexity, accuracy and fluency. *Journal of Second Language Studies*, *3*(1), 31–54. https://doi.org/10.1075/jsls.19025.khe

Kibble, J. D. (2017). Best practices in summative assessment. *Advances in Physiology Education*, *41*(1), 110–119. https://doi.org/10.1152/advan.00116.2016

Krebs, R., Rothstein, B., & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning*, *17*(2), 627–650. *https*://doi.org/10.1007/s11409-022-09302-1

Kuiken, F., & Vedder, I. (2018). *Assessing functional adequacy of L2 performance in a task-based approach* (pp. 265–285). https://doi.org/10.1075/tblt.10.11kui

Kuiken, F., & Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International Journal of Applied Linguistics*, *29*(2), 192–210. https://doi.org/10.1111/ijal.12256

Kuiken, F., & Vedder, I. (2022). Measurement of functional adequacy in different learning contexts. TASK, 2(1), 8–32. https://doi.org/10.1075/task.00013.kui

Kustitskaya, T. A., Kytmanov, A. A., & Noskov, M. V. (2022). Early Student-at-Risk Detection by Current Learning Performance and Learning Behavior Indicators. *Cybernetics and Information Technologies*, *22*(1), 117–133. https://doi.org/10.2478/cait-2022-0008

Kyle, K., & Crossley, S. A. (2014). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, *49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *The Modern Language Journal*, *102*(2), 333–349. https://doi.org/10.1111/modl.12468

Lambert, C., & Nakamura, S. (2018). Proficiency-related variation in syntactic complexity: A study of English L1 and L2 oral descriptive discourse. *International Journal of Applied Linguistics*, *29*(2), 248–264. https://doi.org/10.1111/ijal.12224

Lasagabaster, D. (2022). *English-Medium Instruction in Higher Education*. Cambridge University. https://doi.org/10.1017/9781108903493

Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching*, *52*(4), 524–536. https://doi.org/10.1017/s0261444819000247

Lee, J. (2018). Task Complexity, Cognitive Load, and L1 Speech. *Applied Linguistics*, *40*(3), 506–539. https://doi.org/10.1093/applin/amx054

Lertsakulbunlue, S., & Kantiwong, A. (2024). Development of peer assessment rubrics in simulation-based learning for advanced cardiac life support skills among medical students. *Advances in Simulation*, *9*(1). https://doi.org/10.1186/s41077-024-00301-7

Lewis, D. D. (1992). *An evaluation of phrasal and clustered representations on a text categorization task*. 37–50. https://doi.org/10.1145/133160.133172

Li, D., & Zhang, L. (2021). Contextualizing feedback in L2 writing: the role of teacher scaffolding. *Language Awareness*, *31*(3), 328–350. https://doi.org/10.1080/09658416.2021.1931261

Li, J., & Yuan, F. (2024). Does engagement with feedback matter? Unveiling the impact of learner engagement and grit on EFL learners' English writing achievements. *Language Teaching Research*. https://doi.org/10.1177/13621688241257865

Liao, S. J., Zhang, L. J., & May, S. (2025). English-medium instruction (EMI) language policy and implementation in China's higher education system: growth, challenges, opportunities, solutions, and future directions. *Current Issues in Language Planning*, *ahead-of-print* (ahead-of-print), 1–20. https://doi.org/10.1080/14664208.2025.2453263

Lin, T., & Lei, J. (2021). English-Medium Instruction and Content Learning in Higher Education: Effects of Medium of Instruction, English Proficiency, and Academic Ability. *Sage Open*, *11*(4). https://doi.org/10.1177/21582440211061533

Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X., & Hu, R. (2021). Sense-aware lexical sophistication indices and their relationship to second language writing quality. *Behavior Research Methods*, *54*(3), 1444–1460. https://doi.org/10.3758/s13428-021-01675-6

Luo, X., & Sun, P. P. (2025). Greater complexity and higher efficiency: Effects of translanguaging during collaborative pre-task planning on EFL learners' speaking performance. *Language Teaching Research*. https://doi.org/10.1177/13621688241307029

Macaro, E. (2022). English Medium Instruction: What do we know so far and what do we still need to find out? *Language Teaching*, *55*(4), 533–546. https://doi.org/10.1017/s0261444822000052

Macaro, E., Hultgren, A. K., Kirkpatrick, A., & Lasagabaster, D. (2017). English medium instruction: Global views and countries in focus. *Language Teaching*, *52*(02), 231–248. https://doi.org/10.1017/s0261444816000380

Manchón, R. M., & Sanz, C. (2023). Introduction to the special issue on working memory and L2 writing: Implications for SLA individual differences research. *Studies in Second Language Acquisition*, *45*(3), 599–621. https://doi.org/10.1017/s0272263123000359

Massler, U., Stotz, D., & Queisser, C. (2014). Assessment instruments for primary CLIL: the conceptualisation and evaluation of test tasks. *The Language Learning Journal*, *42*(2), 137–150. https://doi.org/10.1080/09571736.2014.891371

Mccutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*(3), 299–325. https://doi.org/10.1007/bf01464076

Mckinley, J. (2022). *English-Medium Instruction Practices in Higher Education*. Bloomsbury Plc. https://doi.org/10.5040/9781350167889

Mckinley, J., & Rose, H. (2022). English language teaching and English-medium instruction. *Journal of English-Medium Instruction*, *1*(1), 85–104. https://doi.org/10.1075/jemi.21026.mck

Mckinley, J., Rose, H., & Curdt-Christiansen, X. L. (2022). EMI in Chinese higher education: the Muddy water of 'Englishisation.' *Applied Linguistics Review*, *14*(6), 1475–1481. https://doi.org/10.1515/applirev-2022-0015

Mcnamara, D. S., Crossley, S. A., & Mccarthy, P. M. (2009). Linguistic Features of Writing Quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

Mendonça, P. C., Quintal, F., & Mendonça, F. (2025). Evaluating LLMs for Automated Scoring in Formative Assessments. *Applied Sciences*, *15*(5), 2787. https://doi.org/10.3390/app15052787

Mohsen, M. A. (2022). Computer-Mediated Corrective Feedback to Improve L2 Writing Skills: A Meta-Analysis. *Journal of Educational Computing Research*, *60*(5), 1253–1276. https://doi.org/10.1177/07356331211064066

Newman, I., Lim, J., & Pineda, F. (2013). Content Validity Using a Mixed Methods Approach. *Journal of Mixed Methods Research*, *7*(3), 243–260. https://doi.org/10.1177/1558689813476922

Nguyen, H. T. (2018). *English-Medium-Instruction Management: The Missing Piece in the Internationalisation Puzzle of Vietnamese Higher Education* (Vol. 51, pp. 119–137). Springer. https://doi.org/10.1007/978-3-319-78492-2_7

Nguyen, H. T., Hamid, M. O., & Moni, K. (2016). English-medium instruction and self-governance in higher education: the journey of a Vietnamese university through the institutional autonomy regime. *Higher Education*, *72*(5), 669–683. https://doi.org/10.1007/s10734-015-9970-y

Nguyen, H. T., Walkinshaw, I., & Pham, H. H. (2017). *EMI Programs in a Vietnamese University: Language, Pedagogy and Policy Issues* (Vol. 21, pp. 37–52). Springer. https://doi.org/10.1007/978-3-319-51976-0_3

Nguyen, T. H. H., Ho, T. N., Do, T. M. D., & Pham, T. T. T. (2024). Factors Affecting Learner Autonomy in EMI Studying of English-Major Students at some Economics Universities in Hanoi, Vietnam. *International Journal of Language Instruction*, *3*(3), 36–53. https://doi.org/10.54855/ijli.24333

Nguyen, P., & Sercu, L. (2021). Student s ' Perceptions of Content- and Language-Integrated Learning in Vietnam : A Survey Study. *International Journal of Learning, Teaching and Educational Research*, *20*(12), 1–18. https://doi.org/10.26803/ijlter.20.12.1

Norris, J. M., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Otokiti, B. O., Olinmah, F. I., Abiola-Adams, O., Abutu, D. E., Okoli, I., & Imohiosen, C. (2022). Designing Interactive Visual Analytics Frameworks for Higher Education: Feedback and Satisfaction Insights. *International Journal of Social Science Exceptional Research*, *1*(2), 156–163. https://doi.org/10.54660/ijsser.2022.1.2.156-163

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Peng, J.-E., & Xie, X. (Serina). (2021). English-Medium Instruction as a Pedagogical Strategy for the Sustainable Development of EFL Learners in the Chinese Context: A Meta-Analysis of Its Effectiveness. *Sustainability*, *13*(10), 5637. https://doi.org/10.3390/su13105637

Phyak, P. (2023). Producing the disciplined English-speaking subjects: Language policing, development ideology, and English medium of instruction policy. *Language in Society*, *53*(2), 321–343. https://doi.org/10.1017/s0047404523000052

Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, *36*(2), 161–179. https://doi.org/10.1177/0265532216669537

Rahimi, M., & Zhang, L. J. (2018). Writing task complexity, students' motivational beliefs, anxiety and their writing production in English as a second language. *Reading and Writing*, *32*(3), 761–786. https://doi.org/10.1007/s11145-018-9887-9

Rahman, M. M., & Mehar Singh, M. K. (2019). Language ideology of English-medium instruction in higher education. *English Today*, *36*(4), 40–46. https://doi.org/10.1017/s0266078419000294

Rakedzon, T., & Baram-Tsabari, A. (2016). Assessing and improving L2 graduate students' popular science and academic writing in an academic writing course. *Educational Psychology*, *37*(1), 48–66. https://doi.org/10.1080/01443410.2016.1192108

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2016). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

Richards, J. C., & Pun, J. (2021a). A Typology of English-Medium Instruction. *RELC Journal*, *54*(1), 216–240. https://doi.org/10.1177/0033688220968584

Richards, J. C., & Pun, J. (2021b). Teacher strategies in implementing English medium instruction. *ELT Journal*, *76*(2), 227–237. https://doi.org/10.1093/elt/ccab081

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(3). https://doi.org/10.1515/iral.2007.009

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, *52*(2), 161–197. https://doi.org/10.1080/03057267.2016.1206351

Rose, H., Macaro, E., Sahan, K., Aizawa, I., Zhou, S., & Wei, M. (2021). Defining English Medium Instruction: Striving for comparative equivalence. *Language Teaching*, *56*(4), 539–550. https://doi.org/10.1017/s0261444821000483

Sahan, K., Galloway, N., & Mckinley, J. (2022). 'English-only' English medium instruction: Mixed views in Thai and Vietnamese higher education. *Language Teaching Research*, *29*(2), 657–676. https://doi.org/10.1177/13621688211072632

Saricaoglu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL*, *33*(3), 265–277. https://doi.org/10.1017/s0958344021000021

Shen, C., Shi, P., Guo, J., Xu, S., & Tian, J. (2023). From process to product: writing engagement and performance of EFL learners under computer-generated feedback instruction. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1258286

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic Writing Development at the University Level. *Written Communication*, *33*(2), 149–183. https://doi.org/10.1177/0741088316631527

Sugita, Y. (2006). The impact of teachers' comment types on students' revision. *ELT Journal*, *60*(1), 34–41. https://doi.org/10.1093/elt/cci079

Sun, B., & Révész, A. (2021). The Effects of Task Repetition on Child EFL Learners' Oral Performance. *Canadian Journal of Applied Linguistics*, *24*(2), 30–47. https://doi.org/10.37213/cjal.2021.31382

Sykes, M., Rosenberg-Yunger, Z. R. S., Quigley, M., Gupta, L., Thomas, O., Robinson, L., Caulfield, K., Ivers, N., & Alderson, S. (2024). Exploring the content and delivery of feedback facilitation co-interventions: a systematic review. *Implementation Science : IS*, *19*(1). https://doi.org/10.1186/s13012-024-01365-9

T Kellogg, R., P Whiteford, A., E Turner, C., Cahill, M., & Mertens, A. (2013). Working Memory in Written Composition: A Progress Report. *Journal of Writing Research*, *5*(2), 159–190. https://doi.org/10.17239/jowr-2013.05.02.1

Tai, K. W. H. (2023). Shaping student responses into academic expressions: analysing an English medium instruction history classroom from a translanguaging perspective. *International Journal of Bilingual Education and Bilingualism*, *27*(4), 550–580. https://doi.org/10.1080/13670050.2023.2232089

Tai, K. W. H., & Zhao, Y. V. (2022). Success factors for English as a second language university students' attainment in academic English language proficiency: exploring the roles of secondary school medium-of-instruction, motivation and language learning strategies. *Applied Linguistics Review*, *15*(2), 611–641. https://doi.org/10.1515/applirev-2022-0049

Tajik, M. A., Abdul Manan, S., Arvatu, A.-C., & Shegebayev, M. (2022). Growing pains: graduate students grappling with English medium instruction in Kazakhstan. *Asian Englishes*, *26*(1), 249–267. https://doi.org/10.1080/13488678.2022.2119970

Tan, S., Cho, Y. W., & Xu, W. (2022). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments*, *31*(10), 7276–7286. https://doi.org/10.1080/10494820.2022.2066137

Thi, N. K., & Nikolov, M. (2021). How Teacher and Grammarly Feedback Complement One Another in Myanmar EFL Students' Writing. *The Asia-Pacific Education Researcher*, *31*(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Toh, G. (2019). Challenges in English-medium instruction (EMI) at a Japanese university. *World Englishes*, *39*(2), 334–347. https://doi.org/10.1111/weng.12434

Trace, J., Janssen, G., & Meier, V. (2016). Measuring the impact of rater negotiation in writing performance assessment. *Language Testing*, *34*(1), 3–22. https://doi.org/10.1177/0265532215594830

Tran, L. T., & Nguyen, H. T. (2018). *Internationalisation of Higher Education in Vietnam Through English Medium Instruction (EMI): Practices, Tensions and Implications for Local Language Policies* (pp. 91–106). Springer. https://doi.org/10.1007/978-3-319-77655-2_6

Tsao, J.-J. (2021). Effects of EFL Learners' L2 Writing Self-efficacy on Engagement with Written Corrective Feedback. *The Asia-Pacific Education Researcher*, *30*(6), 575–584. https://doi.org/10.1007/s40299-021-00591-9

Vidal, K., & Jarvis, S. (2018). Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research*, *24*(5), 568–587. https://doi.org/10.1177/1362168818817945

Wang, K., Yuan, R., & De Costa, P. I. (2025). A critical review of English medium instruction (EMI) teacher development in higher education: From 2018 to 2022. *Language Teaching*, *58*(2), 141–172. https://doi.org/10.1017/s0261444824000351

Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J. T., & Demark, S. F. (2004). Design Rationale for a Complex Performance Assessment. *International Journal of Testing*, *4*(4), 303–332. https://doi.org/10.1207/s15327574ijt0404_2

Wind, S. A., & Peterson, M. E. (2017). A systematic review of methods for evaluating rating quality in language assessment. *Language Testing*, *35*(2), 161–192. https://doi.org/10.1177/0265532216686999

Wingate, U., Andon, N., & Cogo, A. (2011). Embedding academic writing instruction into subject teaching: A case study. *Active Learning in Higher Education*, *12*(1), 69–81. https://doi.org/10.1177/1469787410387814

Xie, W., & Curle, S. (2020). Success in English Medium Instruction in China: significant indicators and implications. *International Journal of Bilingual Education and Bilingualism*, *25*(2), 585–597. https://doi.org/10.1080/13670050.2019.1703898

Xu, X., Rose, H., Mckinley, J., & Zhou, S. (2021). The incentivisation of English medium instruction in Chinese universities: policy misfires and misalignments. *Applied Linguistics Review*, *14*(6), 1539–1561. https://doi.org/10.1515/applirev-2021-0181

Zedelius, C. M., Mills, C., & Schooler, J. W. (2018). Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods*, *51*(2), 879–894. https://doi.org/10.3758/s13428-018-1137-1

Zhou, S., Fung, D., & Thomas, N. (2023). Towards deeper learning in EMI lectures: the role of English proficiency and motivation in students' deep processing of content

knowledge. *Journal of Multilingual and Multicultural Development*, *46*(7), 1–16. https://doi.org/10.1080/01434632.2023.2248078

Zhou, Y., & Curle, S. (2024). Decoding motivation in English medium instruction: situated expectancy-value theory and student performance at a Chinese transnational university. *Journal of Multilingual and Multicultural Development*, *ahead-of-print*(ahead-of-print), 1–18. https://doi.org/10.1080/01434632.2024.2419967