Original Researcher Article

Value-in-Use in English-Medium Instruction Writing: Interlanguage-Mediated Development Across Genre Touchpoints

Phuong Bao Tran Nguyen[1], Tran Nam Phuong Nguyen[2]
[1]School of Foreign Languages, Can Tho University, Can Tho City, Vietnam
[2]Nam Phuong Center for English, Can Tho City, Vietnam

ABSTRACT

As EMI expands in higher education, academic writing has become a high-stakes site where students must simultaneously meet disciplinary demands and develop L2 writing competence, yet evidence on EMI-related writing gains remains mixed and highly task-dependent. This motivates longitudinal, multidimensional tracking of how interlanguage development is enacted across recurring genre "touchpoints" and time constraints, rather than inferred from single end-point writing samples. This study conceptualizes English-Medium Instruction (EMI) writing development as value-in-use enacted across a semester-long journey of genre touchpoints, with learners' interlanguage functioning as the mechanism through which course resources are translated into observable outcomes. Using a within-subject longitudinal design, 80 EMI undergraduates completed five in-class timed writing tasks (three 20-minute touchpoints and two 45-minute essay touchpoints), yielding 400 texts. Texts were analyzed using a multidimensional outcome space (c-CAF): accuracy (error-free T-unit ratio; errors per 100 words), fluency (words per minute; total words), syntactic complexity (clauses per T-unit; mean length of T-unit), lexical diversity (MTLD), and cohesion (cohesive devices per 100 words). Mixed-effects and clustered inference tested (i) developmental change across the semester, (ii) systematic touchpoint and timing effects (20 vs. 45 minutes), and (iii) persistence versus genre sensitivity in recurrent error tendencies. Results showed robust semester gains in syntactic complexity, lexical diversity, and cohesion, while accuracy and writing rate did not improve monotonically. Timing and genre produced a consistent reconfiguration of the performance profile: 45-minute essays elicited substantially longer texts and higher complexity, lexical diversity, and cohesion, but lower writing rate and a small reduction in error-free production. Error-family analysis revealed persistent interlanguage pressure points (articles/determiners, lexical choice/word form, tense/aspect) alongside clear genre sensitivity: essays reduced article and preposition errors but increased lexical choice, tense/aspect, and sentence-boundary problems. The findings advance a journey-based account of EMI writing as value-in-use and offer actionable implications for designing touchpoint sequences that balance discourse expansion with targeted stabilization of persistent interlanguage constraints

Keywords: English-medium instruction (EMI); value-in-use; customer journey; touchpoints; interlanguage; L2 writing development; complexity–accuracy–fluency (CAF); cohesion; lexical diversity; timed writing; mixed-effects modeling

1. INTRODUCTION:

Higher education is increasingly organized through market-oriented logics in which institutions differentiate programs, bundle support services, and compete on perceived quality and outcomes. Within this environment, the "student-as-consumer" framing has become a visible—if contested—lens for understanding how learners evaluate educational offerings and the experiences they receive (Molesworth et al., 2009). From a consumer research standpoint, this shift is consequential because education resembles a high-involvement service in which benefits are difficult to evaluate prior to consumption and are realized primarily through sustained participation and co-production over time. As a result, value is best conceptualized as value-in-use, emerging as consumers integrate institutional resources with their own capabilities during service consumption (Vargo & Lusch, 2004; Zeithaml, 1988). In educational services, the quality of the "value proposition" therefore depends not only on

institutional branding but also on how learning is designed, delivered, and experienced across a sequence of course encounters.

English-medium instruction (EMI) in higher education provides a particularly informative context for examining value realization in educational services. EMI is frequently positioned as a premium program attribute associated with internationalization, employability, and access to global knowledge. Yet evidence syntheses indicate that EMI implementation quality varies widely and that language-related constraints can shape both learning experiences and outcomes (Macaro et al., 2018). These constraints are likely to be especially pronounced in emerging EMI markets such as Vietnam, where students' preparedness and language support provisions may differ substantially across programs and institutions. In Vietnamese transnational university settings, English has been described as a "double barrier" because it functions simultaneously as the medium through which content must be learned and the criterion on which

achievement is evaluated (Yao et al., 2022). This dual function can increase the sacrifices students must invest—time, effort, anxiety, and cognitive load—to obtain the promised benefits of EMI, rendering the gap between marketing claims and realized outcomes particularly salient.

Service-dominant logic (SDL) provides a coherent theoretical foundation for analyzing EMI as a co-produced service. SDL conceptualizes value as co-created through the integration of provider resources (e.g., curriculum design, genre models, feedback systems, and instructional expertise) and consumer resources (e.g., time, effort, self-regulation, and prior competencies) (Vargo & Lusch, 2004). Complementing SDL, customer-journey research emphasizes that outcomes are shaped across multiple touchpoints rather than a single encounter and that organizations must understand experience and performance as dynamic processes unfolding over time (Lemon & Verhoef, 2016). In EMI writing courses, the "journey" is operationalized through a sequence of writing events—task instructions, timed writing sessions, feedback cycles, and genre progression - each of which imposes distinct demands and creates distinct opportunities for realizing value.

A central premise of the present study is that the key consumer-side mechanism shaping value realization in EMI writing is learners' interlanguage. Interlanguage is a developing linguistic system that is systematic yet adaptive, reflecting what learners can currently mobilize under specific task and time constraints (Selinker, 1972). Conceptualized within SDL, interlanguage can be treated as an individual-level capacity resource that students bring into each service encounter, conditioning how efficiently they can convert provider inputs (instruction, models, feedback) into usable academic output (Vargo & Lusch, 2004). This mechanism is especially important in EMI writing because students must coordinate multiple goals simultaneously: generating ideas, organizing arguments, selecting genre-appropriate language, and maintaining linguistic control. Under time pressure, interlanguage constraints may surface as persistent error patterns, reduced lexical precision, weaker cohesion, or constrained syntactic elaboration; conversely, interlanguage growth should manifest as improved control and more effective deployment of academic discourse resources in increasingly demanding genres (Selinker, 1972).

To evaluate value-in-use empirically, consumer research often relies on perceptions (e.g., satisfaction or perceived value). However, educational contexts can also support behavioral measurement because consumers' outputs are observable artifacts of co-production. In EMI writing, students' written texts provide auditable traces of realized benefits and experienced constraints, offering an outcome-focused complement to self-report evaluations (Lemon & Verhoef, 2016; Zeithaml, 1988). A robust operationalization should capture multidimensional writing outcomes rather than a single global score. In second language performance research, the complexity–accuracy–fluency (CAF) framework has been widely used to model proficiency as multi-componential and sensitive to task demands, including potential trade-offs under constrained attentional resources (Housen et al., 2012; Norris & Ortega, 2009). In writing, cohesion is also crucial because textual connectivity strongly influences comprehensibility and evaluations of quality; cohesion theory specifies the linguistic resources that link clauses and sentences into unified discourse (Halliday & Hasan, 1976). Taken together, these perspectives support a combined c-CAF approach (cohesion + CAF) to represent objective outcome profiles associated with EMI writing development under realistic classroom constraints.

Despite growing EMI scholarship, two limitations remain salient for outcome-focused analyses. First, many EMI studies emphasize stakeholder perceptions and implementation challenges, whereas fewer studies trace within-student writing development longitudinally using repeated structured tasks that allow researchers to separate growth from task effects (Macaro et al., 2018; Yao et al., 2022). Second, even where writing is assessed, outcome measures and task designs vary substantially, complicating inference about whether observed changes reflect interlanguage development or shifts in task/genre demands. Methodological guidance in CAF research therefore stresses alignment between theory, task conditions, and measurement choices when interpreting development from repeated performance (Norris & Ortega, 2009). Genre-based writing theory reinforces this point by emphasizing that writing competence involves control over socially recognized text types; different genres systematically elicit different rhetorical moves and linguistic resources, meaning that touchpoint design (task type and sequencing) is integral to the outcomes learners can realize (Hyland, 2007).

Against this backdrop, the present study examines value realization in an EMI writing course by tracking students' written outcomes across a structured sequence of five timed tasks spanning correspondence genres and academic essays. The design is longitudinal and within-subject, enabling analysis of (a) developmental change across the semester and (b) systematic differences attributable to genre-based touchpoints under timed delivery. Specifically, the task sequence comprises an informal letter, a formal email, an opinion paragraph, an advantages–disadvantages essay, and an argumentative essay. In line with classroom delivery constraints, correspondence tasks are completed in 20 minutes and essay tasks in 45 minutes. Outcomes are operationalized via c-CAF indicators, including accuracy measures with established validity in L2 writing research (e.g., error-based indices) (Polio, 1997), fluency indices anchored to production constraints, syntactic complexity indicators consistent with CAF measurement practices (Norris & Ortega, 2009), lexical diversity indicators where appropriate (Housen et al., 2012), and cohesion indices grounded in cohesion theory (Halliday & Hasan, 1976). This combination allows the study to model how interlanguage-mediated capacity interacts with touchpoint demands, producing measurable outcome profiles across the journey.

The study addresses three research questions:

RQ1: How do students' interlanguage-mediated writing outcomes (accuracy, fluency, syntactic complexity,

lexical diversity, and cohesion) change across the semester?

RQ2: To what extent do task types/genres (course touchpoints) systematically shape the writing outcome profile under timed conditions (20 vs. 45 minutes)?

RQ3: Which interlanguage patterns (e.g., recurrent error tendencies) persist across tasks, and which are most sensitive to genre demands?

By embedding interlanguage as the explanatory mechanism within a service co-production framing, the study contributes to outcome-oriented consumer research on educational services in three ways. First, it operationalises value-in-use through objective behavioural outcomes rather than relying solely on perceptions (Lemon & Verhoef, 2016; Zeithaml, 1988). Second, it models the educational customer journey as a sequence of genre-specific touchpoints that vary in demand and time pressure, clarifying how service design can reshape realised benefits across encounters (Hyland, 2007; Lemon & Verhoef, 2016). Third, it provides a mechanism-based account of heterogeneity in value realisation by treating interlanguage as a dynamic consumer capability resource that conditions co-production effectiveness across time and tasks (Selinker, 1972; Vargo & Lusch, 2004).

2. LITERATURE REVIEW

2.1. From "value" to value-in-use in learning services

This study frames EMI academic writing as a service experience in which learning value is realized in use - that is, while students perform meaningful writing actions under real course constraints—rather than being fully captured by end-of-course grades alone. In service-dominant logic (SDL), value is not embedded in a delivered "product"; it emerges through actors' resource integration and experience over time (Vargo & Lusch, 2004). Translating this into an EMI writing course, students integrate linguistic resources (grammar, lexis, cohesion devices), disciplinary resources (task expectations, genre norms), and situational resources (time limits, platform affordances) while producing texts. The implication is that "learning outcomes" should be examined as trajectory-shaped performance that becomes visible at moments of use—i.e., at repeated course writing events that demand language-in-action, not language-in-isolation.

In consumer and service research, perceived value has long been conceptualized as a trade-off between what is received and what is given (e.g., benefits relative to costs), providing a basis for operationalising value perceptions and value formation (Zeithaml, 1988). In EMI writing, the "costs" can include cognitive load, time pressure, and linguistic risk, whereas the "benefits" can include communicative success, genre control, and increasing linguistic stability. This positions timed academic writing as an analytically rich context for tracing how students "cash out" learning value through performances that are simultaneously cognitive, social, and textual.

2.2. Learning as a journey of touchpoints

Customer experience scholarship has demonstrated that experience is best understood across the customer journey, where value formation is shaped by multiple interactions, or "touchpoints," rather than by a single transaction (Lemon & Verhoef, 2016). Service design operationalises this idea by mapping interactions across interfaces and episodes, emphasising that different touchpoints impose different requirements and therefore can produce different experience and performance outcomes (Patrício et al., 2008). Experience-centric service design further highlights that organisations can intentionally design sequences of touchpoints to create distinctive patterns of engagement and outcomes (Zomerdijk & Voss, 2010).

Adapting these ideas to EMI writing, the semester can be treated as a learning journey composed of recurrent writing touchpoints (e.g., short professional emails vs. longer essays). Each touchpoint instantiates a distinct set of constraints and success criteria (audience, purpose, register, rhetorical moves, and time), meaning it is plausible—indeed likely—that students' writing profiles will vary systematically by task type. This touchpoint view motivates the study's second research question: whether genres (course touchpoints) systematically shape the writing outcome profile under timed conditions.

2.3. EMI as a high-growth context with persistent language-related challenges

EMI has expanded rapidly worldwide, prompting sustained concerns about how language-related demands interact with content learning and assessment practices. Early global mapping work described EMI as a "fast-moving" shift in many systems and highlighted recurrent implementation issues, including student language readiness and pedagogical support (Dearden, 2014). Systematic evidence syntheses similarly emphasize that the research base has grown quickly, while also noting persistent uncertainty and variability in language and learning outcomes across contexts (Macaro et al., 2018). More recently, an updated systematic review of EMI research (covering 2016–2023) underscores both the exponential growth of EMI studies and continuing attention to language-related challenges in higher education EMI settings (Rose et al., 2026). Collectively, this literature justifies studying writing development inside EMI as an authentic, high-stakes setting where language performance is continuously demanded and consequential.

To conceptualize EMI in multilingual universities, the ROAD-MAPPING framework has been proposed as an integrative lens for analyzing the intersecting dimensions of English-medium education, including the roles of languages, agents, practices, and policies (Dafouz & Smit, 2020). In Vietnam specifically, ROAD-MAPPING has been discussed as a useful framework for evaluating EMI implementation in higher education and for foregrounding how contextual conditions shape outcomes (Võ, 2025). For writing development research, the practical implication is that "EMI writing progress" should not be assumed to be uniform; it is more plausibly context- and task-sensitive, varying by the types of writing events that

students must repeatedly complete under institutional constraints.

2.4. Interlanguage as the mechanism linking experience to development

While SDL and touchpoint perspectives help theorize where learning value is realized (in repeated use across course events), interlanguage theory helps explain how development manifests in learner language. Interlanguage refers to the evolving, systematic linguistic system developed by L2 learners that is neither the L1 nor the target language but has its own rule-governed regularities and transitional forms (Selinker, 1972). From this perspective, repeated writing tasks provide repeated observations of a developing system: learners may show progress (greater stability and target-like use), persistence (recurrent, systematic error tendencies), and task-conditioned variability (different outcomes when task demands shift).

This framing aligns directly with the study's first and third research questions. RQ1 treats semester-long changes in writing outcomes as interlanguage-mediated development—observable as shifts in accuracy, fluency, syntactic complexity, lexical diversity, and cohesion across time. RQ3 extends beyond aggregate change to focus on patterns: which tendencies persist across tasks (suggesting entrenched interlanguage routines) and which are sensitive to genre demands (suggesting adaptive, task-contingent deployment of resources).

2.5. Outcomes as c-CAF: complexity, accuracy, fluency, and cohesion

To operationalize interlanguage development in written production, the study adopts a c-CAF outcome space (syntactic/lexical complexity, accuracy, fluency, plus cohesion). CAF has been widely used to describe multidimensional L2 performance and to capture trade-offs that emerge when learners allocate limited attentional resources (Housen et al., 2012). Methodologically, CAF scholarship also cautions that constructs and measures must be chosen with attention to task conditions and interpretive logic, rather than assuming any single index can stand for "proficiency" in general (Norris & Ortega, 2009).

In L2 writing research, accuracy has been operationalized through multiple measures, including error-free units and error density, with careful attention to reliability and the implications of different coding choices (Polio, 1997). In an interlanguage frame, persistent accuracy patterns can be interpreted as stable properties of the learner system—especially when they recur across tasks—whereas accuracy improvements across repeated tasks can signal increasing control or more efficient monitoring under time pressure.

Complexity is typically approached as syntactic and lexical elaboration (Housen et al., 2012). However, complexity can shift with genre demands: some genres prioritize concise functional phrasing, others encourage elaborated clause structure or stance marking. Consistent with CAF cautions, interpreting complexity requires anchoring claims to the tasks being performed and the resources those tasks elicit (Norris & Ortega, 2009).

In timed writing, fluency is often treated as production volume and/or rate—observable as how much text is produced within a constrained time window. Under touchpoint logic, fluency is expected to vary by task length, rhetorical complexity, and planning demands, making it a sensitive indicator of time pressure effects across 20-minute vs. 45-minute conditions.

Cohesion provides the textual "glue" that links ideas and supports reader interpretation. Classic discourse accounts define cohesion as a set of linguistic resources for constructing textual connectedness (Halliday & Hasan, 1976). Computational tools such as Coh-Metrix were developed to automate analyses of cohesion and related discourse features at scale (Graesser et al., 2004). In L2 writing specifically, cohesion features have been shown to develop over a semester and to relate to judgments of writing quality, with distinctions between local, global, and text-level cohesion offering a nuanced view of how learners manage discourse constraints (Crossley et al., 2016). In the present study, cohesion is therefore treated as an essential part of writing value-in-use: it captures whether texts "work" for readers under the communicative conditions imposed by each touchpoint.

2.6. Genre demands as structured variation in the learning journey

A touchpoint-based design treats genres as **structured contexts** that shape what counts as success and what linguistic resources are functionally useful. Genre theory in applied linguistics emphasizes that genres are socially recognized ways of using language to achieve recurring purposes in particular contexts, and learning genres involves learning how to control rhetorical and linguistic choices under those contexts (Hyland, 2004). In academic settings, genre analysis has similarly highlighted that communicative purposes and discourse communities shape text organization and language choices (Swales, 1990). For EMI writing, this means that a "short email" touchpoint is not merely a shorter essay; it is a different communicative situation with different expectations for concision, politeness strategies, information packaging, and cohesion signaling. Accordingly, genre should not only influence mean outcomes; it should plausibly shift the *profile* of c-CAF outcomes across tasks.

2.7. Conceptual framework

Bringing these strands together, the study proposes a compact conceptual chain. First, SDL positions learning value as value-in-use: value becomes visible when students enact writing under authentic constraints rather than when they merely "possess" knowledge (Vargo & Lusch, 2004; Zeithaml, 1988). Second, the semester is treated as a journey composed of multiple writing touchpoints, where each touchpoint constitutes a distinct interaction episode with its own demands (Lemon & Verhoef, 2016; Patrício et al., 2008; Zomerdijk & Voss, 2010). Third, interlanguage theory specifies the developmental mechanism: repeated touchpoints repeatedly sample an evolving learner system, revealing both systematic tendencies and task-conditioned variability (Selinker, 1972). Finally, development is operationalized through c-CAF outcomes—accuracy, fluency, syntactic complexity, lexical diversity, and

cohesion—because these dimensions capture complementary aspects of writing performance and allow testing whether changes reflect improvement, trade-offs, or genre-conditioned adaptation (Housen et al., 2012; Norris & Ortega, 2009; Polio, 1997; Crossley et al., 2016)
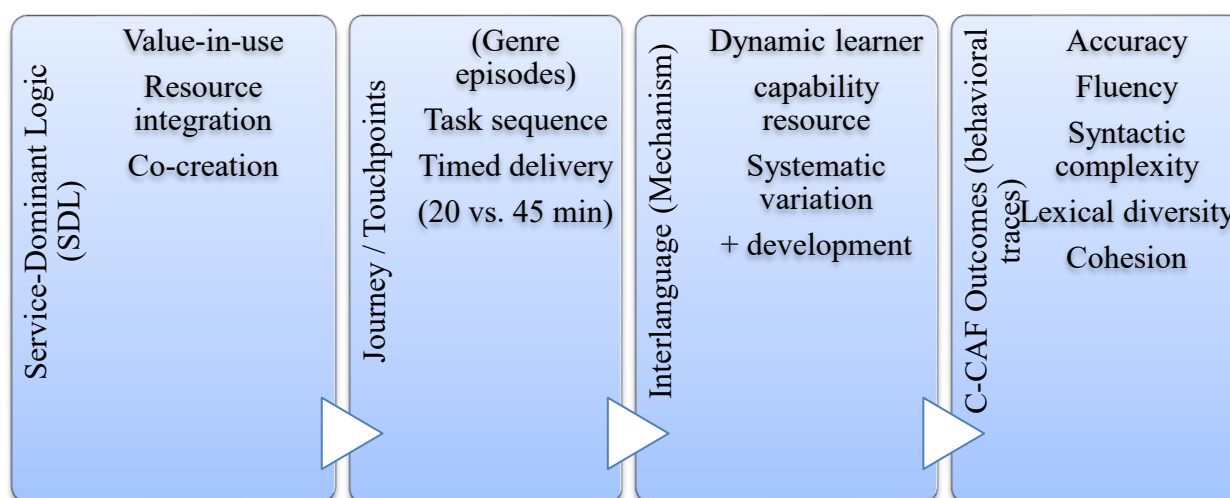

Figure 1-1Conceptual Framework: SDL Journey/Touchpoints Interlanguage c-CAF Outcomes

This framework implies a design that repeatedly observes the same learners across multiple touchpoints so that "value-in-use" and interlanguage development can be traced as within-student change, not merely between-student differences. It also implies that touchpoints should include contrasting genre conditions and realistic time constraints, because time pressure is part of how writing value is produced in classrooms (e.g., short in-class emails versus longer essays) and because CAF theory warns that performance profiles depend on task conditions and attentional allocation (Housen et al., 2012; Norris & Ortega, 2009). Consequently, the study's methodological choices—repeated timed tasks (20 vs. 45 minutes) and within-subject modeling—follow directly from the literature's claim that both experience and development are best understood as trajectories across structured interaction episodes rather than as single-shot outcomes (Lemon & Verhoef, 2016; Selinker, 1972; Macaro et al., 2018).

3. METHODOLOGY

3.1. Research design

This study employed a semester-long, within-subject longitudinal design using repeated, timed writing tasks that represent course "journey touchpoints" (i.e., genre-based episodes of value-in-use where learners mobilize linguistic resources to accomplish situated communicative goals). The design aligns with service research views of experience unfolding across touchpoints (Lemon & Verhoef, 2016) while treating learner language

as an evolving interlanguage system that develops and varies systematically under changing task demands (Selinker, 1972). The primary outcomes were calibrated complexity–accuracy–fluency indicators, complemented by cohesion and lexical diversity measures (hereafter c-CAF), consistent with calls for construct-sensitive operationalization in CAF research (Norris & Ortega, 2009) and multi-dimensional proficiency perspectives (Housen et al., 2012). The repeated-measures structure enables estimation of (a) developmental change across the semester (RQ1), (b) systematic task/genre and time-limit effects (RQ2), and (c) persistence versus genre-sensitivity in recurrent interlanguage patterns (RQ3).

3.2. Research context

The study was conducted in an English-Medium Instruction (EMI) undergraduate program at a regional Vietnamese public university, where disciplinary learning and assessment are delivered primarily through English and where students regularly complete academic writing tasks as part of course requirements. EMI contexts are widely reported to create sustained linguistic demands and uneven language-related challenges for students, particularly in academically literate production (Macaro et al., 2018). In the focal writing course, genre tasks were intentionally sequenced to mirror authentic academic communication touchpoints (e.g., brief functional messages and longer argumentative/academic texts). All tasks were completed in class under controlled timing to minimize confounding from outside resources and to foreground time pressure as a task-processing condition relevant to performance trade-offs (Polio et al., 1998).

3.3. Participants and sampling technique

Participants were 80 undergraduate students enrolled in the same EMI writing course during one semester. A convenience sampling approach was used because the cohort constituted an intact instructional group; however, analytic inclusion required completion of at least 80% of writing touchpoints and provision of informed consent for research use of coursework artifacts. This approach supports internal comparability by holding instructional exposure constant while enabling robust within-subject estimation of change and task effects. Participant identifiers were replaced with coded IDs prior to analysis, and no individually identifying information was retained in the analytic dataset.

3.4. Procedure and data analysis

Procedure.

Writing was elicited at multiple points across the semester using a fixed set of course touchpoints operationalized as distinct genres (e.g., functional email writing and longer academic essays). Tasks were administered under timed conditions designed to reflect two common classroom constraints: brief writing (20 minutes; e.g., email or short functional genre) versus extended writing (45 minutes; e.g., essay-length genres). Each participant completed all scheduled touchpoints, producing a repeated corpus suitable for modeling developmental trajectories (RQ1) and within-student sensitivity to genre and time limits (RQ2). Prompts were aligned to course objectives and administered under standardized instructions; the same administration procedures (timing, setting, allowable materials) were used for all students to improve comparability across tasks.

Outcome measures (c-CAF and cohesion)

For each text, c-CAF indicators were computed to represent performance across accuracy, fluency, syntactic complexity, lexical diversity, and cohesion. Accuracy was operationalized using established written accuracy metrics (e.g., error-based ratios such as errors per T-unit/100 words and/or proportion of error-free units), following recommendations that accuracy measures be explicit, reliable, and fit-for-purpose in L2 writing research (Polio, 1997). Complexity and fluency indices were derived from sentence/clause and production-length proxies commonly used in CAF work, with attention to measurement sustainability and construct validity (Norris & Ortega, 2009). Lexical diversity was indexed using robust measures designed to mitigate text-length sensitivity, such as MTLD (McCarthy & Jarvis, 2010). Cohesion was operationalized via computational indices capturing connective use and related cohesion features (Graesser et al., 2004), consistent with evidence that cohesive features can develop across a semester and relate to genre demands and quality judgments in L2 writing (Crossley et al., 2016).

Interlanguage pattern coding (RQ3).

To address persistence and genre-sensitivity in interlanguage patterns, recurrent error tendencies were coded using an error taxonomy (e.g., article use, verb tense/aspect, subject–verb agreement, prepositions, word form, sentence boundary issues). This analytic layer treats errors as observable traces of an evolving interlanguage

system rather than isolated deficits (Selinker, 1972). A subset of scripts (e.g., 20%) was double-coded by trained raters to establish reliability; disagreements were resolved through adjudication and codebook refinement. Recurrent patterns were defined as error categories that (a) appeared in a substantial proportion of students and (b) recurred across multiple touchpoints, while "genre-sensitive" patterns were those whose rates changed materially across genres and/or under different time limits.

Statistical modeling.

Given the repeated-measures structure, primary inference relied on mixed-effects models, which are well suited to longitudinal linguistic outcomes with observations nested within individuals (Baayen et al., 2008). For continuous outcomes (e.g., lexical diversity, cohesion indices, syntactic complexity measures), linear mixed-effects models estimated fixed effects of time (semester progression), task/genre, time limit (20 vs. 45 minutes), and their interactions, with random intercepts for students and (where supported) random slopes for within-student predictors to reduce anti-conservative inference (Barr, 2013). For count or categorical outcomes (e.g., presence/absence of a specific error type; error counts), generalized mixed models with appropriate link functions were used, consistent with guidance to move beyond ANOVA-style treatment of non-normal outcomes (Jaeger, 2008). RQ1 was tested primarily via the fixed effect of time (and, when theoretically motivated, non-linear time terms), RQ2 via fixed effects of genre and time limit and their interaction, and RQ3 via models that predicted error-category rates as a function of genre and time limit while tracking cross-task persistence. Model diagnostics included inspection of residual patterns for continuous outcomes and checks for overdispersion and convergence for generalized models.

3.5. Research ethics

The study followed core human-subjects principles of respect for persons, beneficence, and justice (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Participation was voluntary and based on informed consent; students were informed that opting out would not affect course grades or standing. To minimize coercion in a classroom setting, research consent and data processing were separated from instructional grading: texts were graded as usual for course purposes, then de-identified and exported for research analysis after grade finalization. All data were stored on password-protected devices with access limited to the research team, and reporting was conducted at aggregate level to prevent identification of individual students.

4. RESULTS

4.1. Descriptive results and corpus characteristics

The analytic corpus comprised 80 EMI undergraduates who each completed five in-class timed writing touchpoints, yielding 400 texts in total. The touchpoints included three 20-minute tasks (Informal Letter, Formal Email, Opinion Paragraph) and two 45-minute tasks (Advantages–Disadvantages Essay, Argumentative Essay). Descriptive statistics for all outcomes by

touchpoint are reported in Table 1 (touchpoint-level descriptives) and Table 1a (pooled descriptives across the full corpus). Across all tasks, students produced texts with an overall mean error-free T-unit ratio (EFTU) of 0.416 and 12.00 errors per 100 words, indicating moderate accuracy with substantial interlanguage variability. Mean writing rate across the corpus was 9.29 words per minute (WPM), while linguistic elaboration indicators showed average 1.36 clauses per T-unit and MLTU = 11.47 words, suggesting a developing but still constrained ability to package propositions into extended units. Lexical diversity was moderate (MTLD = 58.30), and cohesion signaling averaged 12.01 cohesive devices per 100 words, with considerable differences by genre and time condition (Table 1). These descriptives provide the baseline profile against which longitudinal change (RQ1), touchpoint/timing effects (RQ2), and recurrent interlanguage patterns (RQ3) were evaluated.

Table 1. Descriptive statistics for c-CAF outcomes by touchpoint (N = 80 per touchpoint)

Outcome	Informal Letter (20m)	Formal Email (20m)	Opinion Paragraph (20m)	Adv – Disadv Essay (45m)	Argumentative Essay (45m)
Words	170.04 (32.01)	185.70 (30.25)	223.51 (33.40)	364.18 (50.23)	423.30 (51.47)
WPM	8.50 (1.60)	9.29 (1.51)	11.18 (1.67)	8.09 (1.12)	9.41 (1.14)
EFTU ratio	0.392 (0.082)	0.451 (0.083)	0.430 (0.089)	0.398 (0.080)	0.411 (0.078)
Errors /100 words	12.53 (2.49)	12.17 (2.11)	10.83 (2.54)	11.94 (2.36)	12.55 (1.98)
Clauses/T-unit	1.12 (0.09)	1.18 (0.09)	1.33 (0.07)	1.53 (0.09)	1.63 (0.08)
MLTU	9.48 (1.08)	10.05 (1.07)	11.43 (1.07)	12.87 (1.07)	13.54 (0.99)
MTLD	48.53 (9.19)	52.75 (7.78)	56.41 (8.94)	65.22 (9.47)	68.56 (8.06)
Cohesion devices /100 words	9.26 (2.01)	9.93 (2.03)	11.11 (2.03)	14.42 (2.08)	15.31 (2.14)

4.2. RQ1: Longitudinal change in interlanguage-mediated outcomes across the semester

To address RQ1, models were fitted to test whether students' c-CAF outcomes exhibited systematic change across the ordered touchpoints (from earlier to later in the semester). Results indicate clear developmental gains in syntactic complexity, lexical diversity, and cohesion, but comparatively stable patterns in accuracy and fluency.

First, syntactic complexity increased significantly over time. The timepoint slope was positive for clauses per T-unit (b = 0.136 per touchpoint, 95% CI [0.130, 0.142], p < .001) and for mean length of T-unit (MLTU) (b = 1.094, 95% CI [1.010, 1.179], p < .001), indicating that students increasingly produced more elaborated clause packaging and longer T-units as the semester progressed. Second, lexical diversity increased strongly, with MTLD rising by approximately 5.25 units per touchpoint (b = 5.253, 95% CI [4.673, 5.832], p < .001). Third, cohesion signaling increased substantially, with cohesive devices per 100 words increasing by 1.66 per touchpoint (b = 1.660, 95% CI [1.520, 1.799], p < .001). Descriptively, from the first to the final touchpoint, students showed sizeable increases (+0.51 clauses/T-unit, +4.06 MLTU, +20.03 MTLD, +6.05 cohesion devices/100 words), consistent with a semester-long trend toward richer lexical deployment and more explicit textual connectivity (Table 1).

By contrast, accuracy did not show a monotonic improvement across the semester. Neither the EFTU ratio (b = −0.001, 95% CI [−0.006, 0.003], p = .490) nor errors per 100 words (b = −0.021, 95% CI [−0.168, 0.126], p = .779) exhibited a significant linear time trend. Descriptive patterns suggested localized peaks (e.g., relatively high EFTU at the Formal Email touchpoint), but these did not translate into an overall upward trajectory through the essay touchpoints (Table 1). Finally, fluency (WPM) also did not increase linearly (b = 0.062, 95% CI [−0.024, 0.147], p = .158), reflecting task-contingent modulation of production speed rather than steady acceleration across timepoints.

Taken together, the RQ1 findings indicate that across the semester, students' interlanguage-mediated development was most visible in dimensions associated with linguistic elaboration and discourse organization (complexity, lexical diversity, cohesion), while accuracy and speed appeared comparatively constrained and sensitive to shifting genre demands.

4.3. RQ2: Touchpoint and timing effects on c-CAF outcome profiles (20 vs. 45 minutes)

RQ2 examined whether the writing outcome profile varied systematically by touchpoint (genre) and whether the timed condition (20 vs. 45 minutes) shifted outcomes in consistent ways. Results show strong touchpoint

structuring of performance and a coherent timing-related reconfiguration of the c-CAF profile.

At the touchpoint level, outcomes differed reliably across the five genres (joint tests for touchpoint effects were significant for all outcomes; detailed coefficients are reported in Table 2). Descriptively, the 20-minute touchpoints tended to elicit relatively higher production rate and, in specific tasks, comparatively favorable accuracy indicators. For example, the Opinion Paragraph yielded the highest mean WPM (11.18) and the lowest mean error density (10.83 errors/100 words), while the Formal Email showed the highest mean EFTU ratio (0.451) (Table 1). In contrast, the 45-minute essay touchpoints elicited a markedly different profile, characterized by higher elaboration and discourse-level signaling. Both essays showed higher mean clauses per T-unit and cohesion rates, with the Argumentative Essay producing the highest observed complexity (clauses/T-unit = 1.63) and cohesion (15.31 devices/100 words) (Table 1). These differences demonstrate that "genre touchpoints" functioned as structured performance contexts, shaping which aspects of interlanguage were most visible and most demanded.

When tasks were grouped by timed condition, the contrast sharpened further. Relative to the 20-minute condition, 45-minute tasks produced substantially longer texts (difference b = +200.65 words, 95% CI [192.99, 208.31], $p < .001$), yet writing rate decreased (b = −0.904 WPM, 95% CI [−1.132, −0.677], $p < .001$), suggesting that longer tasks encouraged greater planning and revision effort per unit time rather than faster output. Consistent with this, 45-minute tasks showed strong increases in complexity (clauses/T-unit b = +0.368, $p < .001$; MLTU b = +2.886, $p < .001$), lexical diversity (MTLD b = +14.329, $p < .001$), and cohesion (devices/100 words b = +4.767, $p < .001$) (Table 2). Accuracy exhibited a small but statistically reliable decline in the 45-minute condition for EFTU (b = −0.0197, $p < .001$), while changes in error density were comparatively modest (errors/100 words b = +0.397, $p = .071$). Overall, the timing results indicate that extended writing conditions fostered more elaborated and cohesive discourse with broader lexical deployment, but with slightly weaker error-free production, consistent with a reallocation of attentional resources under more demanding genre conditions.

Table 2. Mixed-effects model summary for RQ1 and RQ2 (fixed effects; clustered by student)

Outcome	Predictor	b	95% CI	p
Clauses/T-unit	Timepoint (RQ1)	0.136	[0.130, 0.142]	<.001
MLTU	Timepoint (RQ1)	1.094	[1.010, 1.179]	<.001
MTLD	Timepoint (RQ1)	5.253	[4.673, 5.832]	<.001
Cohesion/100w	Timepoint (RQ1)	1.660	[1.520, 1.799]	<.001
EFTU ratio	Timepoint (RQ1)	−0.001	[−0.006, 0.003]	.490
Errors/100w	Timepoint (RQ1)	−0.021	[−0.168, 0.126]	.779
WPM	Timepoint (RQ1)	0.062	[−0.024, 0.147]	.158
Words	45 min vs 20 min (RQ2)	200.65	[192.99, 208.31]	<.001
WPM	45 min vs 20 min (RQ2)	−0.904	[−1.132, −0.677]	<.001
Clauses/T-unit	45 min vs 20 min (RQ2)	0.368	[0.351, 0.384]	<.001
MLTU	45 min vs 20 min (RQ2)	2.886	[2.653, 3.119]	<.001
MTLD	45 min vs 20 min (RQ2)	14.329	[12.811, 15.847]	<.001
Cohesion/100w	45 min vs 20 min (RQ2)	4.767	[4.381, 5.153]	<.001
EFTU ratio	45 min vs 20 min (RQ2)	−0.0197	[−0.0306, −0.0088]	<.001
Errors/100w	45 min vs 20 min (RQ2)	0.397	[−0.034, 0.827]	.071

4.4. RQ3: Persistent versus genre-sensitive interlanguage error patterns

RQ3 investigated which recurrent interlanguage tendencies persisted across tasks and which were most sensitive to genre demands. Across the full corpus, error-category descriptives (Table 3) revealed a stable set of high-frequency pressure points: articles/determiners (mean 2.61 per 100 words), word form/lexical choice (2.40), and verb tense/aspect (2.22), followed by prepositions (2.01), subject–verb agreement (1.44), and sentence boundary/punctuation (1.32). This distribution suggests that students' interlanguage constraints were

concentrated in function-word systems and lexico-grammatical encoding—areas widely known to be persistent for L2 writers in academically demanding contexts.

At the same time, the error profile displayed systematic genre sensitivity, particularly when comparing essay touchpoints (45 minutes) to shorter touchpoints (20 minutes). Essays showed lower article/determiner error rates (difference b = −0.481 per 100 words, p < .001) and lower preposition error rates (b = −0.215, p = .011), but higher word form/lexical choice errors (b = +0.422, p < .001), higher tense/aspect errors (b = +0.348, p < .001), and more sentence boundary issues (b = +0.213, p = .003) (Table 3). These shifts suggest that essay writing amplified pressures on lexical precision, morphological selection, and clause linkage, while some shorter-genre function-word errors diminished. In other words, interlanguage "signatures" were persistent at the cohort level, but their expression was modulated by genre-specific rhetorical and syntactic demands.

At the individual level, rank-order stability in error profiles across adjacent touchpoints was modest (median Spearman $\rho \approx 0.33$), indicating that many students carried forward recognizable interlanguage tendencies while still showing meaningful task-conditioned fluctuations (Table 3 notes). Collectively, the RQ3 findings support a dual characterization of interlanguage: it is sufficiently systematic to generate recurring error families, yet sufficiently adaptive that genre touchpoints systematically shift which constraints are most salient.

Table 3. Error-category descriptives and genre sensitivity (RQ3)

Rates are errors per 100 words. The coefficient b represents the mean difference (45-minute essays − 20-minute tasks) estimated with cluster-robust SEs by student.

Error category (per 100 words)	Overall M	20-min tasks M	45-min tasks M	b (45 min–20 m)	95% CI	p
Articles/determiners	2.61	2.79	2.33	-0.458	[-0.550, -0.366]	<.001
Word form/lexical choice	2.40	2.22	2.67	0.445	[0.356, 0.533]	<.001
Verb tense/aspect	2.22	2.07	2.44	0.369	[0.288, 0.450]	<.001
Prepositions	2.01	2.09	1.89	-0.197	[-0.268, -0.126]	<.001
Subject–verb agreement	1.44	1.44	1.45	0.013	[-0.038, 0.065]	.619
Sentence boundary/punctuation	1.32	1.23	1.46	0.225	[0.177, 0.274]	<.001

To address RQ3, recurrent interlanguage tendencies were examined at the level of error families (errors per 100 words) across all touchpoints and then contrasted between the 20-minute tasks and the 45-minute essay touchpoints. As shown in Table 3, several error families were persistent in the sense that they remained among the most frequent constraints across the full corpus, regardless of task type. In particular, articles/determiners (overall M = 2.61), word form/lexical choice (M = 2.40), and verb tense/aspect (M = 2.22) constituted the dominant pressure points across tasks, followed by prepositions (M = 2.01). This distribution suggests stable interlanguage vulnerabilities in function-word systems and lexico-grammatical encoding that continue to surface across multiple genres, even as learners progress through the semester. At the same time, the error profile was systematically genre-sensitive. Relative to the 20-minute touchpoints, 45-minute essay writing produced a marked reconfiguration: article/determiner errors decreased (b = −0.458, 95% CI [−0.550, −0.366], p < .001) and preposition errors decreased (b = −0.197, 95% CI [−0.268, −0.126], p < .001), whereas word form/lexical choice errors increased (b = +0.445, 95% CI [0.356, 0.533], p < .001), tense/aspect errors increased (b = +0.369, 95% CI [0.288, 0.450], p < .001), and sentence boundary/punctuation problems increased (b = +0.225, 95% CI [0.177, 0.274], p < .001) (Table 3). In contrast, subject–verb agreement exhibited no reliable difference between timing/genre conditions (b = +0.013, p = .619), indicating a comparatively stable error rate that was less sensitive to genre demands. Overall, these findings support a dual characterization of interlanguage in EMI writing: learners display persistent, high-frequency constraints that recur across tasks, while the salience and magnitude of specific error families shift predictably with genre and extended discourse demands in essay touchpoints.

4. DISCUSSION

This study conceptualized EMI writing as value-in-use realized across a semester-long sequence of genre touchpoints, with interlanguage functioning as the capability mechanism through which instructional

resources are converted into observable c-CAF outcomes (accuracy, fluency, syntactic complexity, lexical diversity, and cohesion).

4.1 Interlanguage as Value-in-Use Across EMI Writing Touchpoints: Integrating Development, Task Effects, and Contributions

Regarding RQ1, the dominant developmental signal across the semester was a consistent increase in syntactic complexity, lexical diversity, and cohesion, while accuracy and fluency (WPM) were comparatively stable. This pattern is compatible with accounts of L2 performance in which development may be expressed first as expansion of available linguistic and discourse resources (complexity/lexis/cohesion) before accuracy becomes uniformly more target-like, especially when tasks evolve in rhetorical demands over time (Norris & Ortega, 2009; Skehan, 2009). From an interlanguage perspective, the results align with the view that learner language is systematic yet variable; learners can demonstrate growth in expressive range while maintaining persistent pressure points in form–function mapping (Selinker, 1972). The observed cohesion growth further suggests that development was not purely grammatical, but increasingly reader-oriented and discourse-functional, consistent with cohesion theory and computational discourse work emphasizing the role of cohesion devices in constructing comprehensible academic texts (Halliday & Hasan, 1976; Graesser et al., 2004). Interpreted through SDL, these longitudinal gains indicate that learners increasingly realized "value-in-use" by producing texts that mobilized richer linguistic resources and more explicit textual connectivity under classroom constraints (Vargo & Lusch, 2004).

RQ2 showed that writing outcomes were strongly structured by touchpoint type and timed condition. The 45-minute essay touchpoints elicited substantially greater production volume and higher complexity, lexical diversity, and cohesion, while writing rate decreased and error-free production declined modestly. These results are consistent with models of writing as a set of interacting processes (planning, translating, reviewing), where extended time windows encourage more elaborated discourse construction but also invite more lexical and syntactic risk-taking that can attenuate accuracy (Hayes, 2012). The profile also accords with CAF-oriented reasoning that performance reflects attentional allocation and potential trade-offs across dimensions as task demands increase (Norris & Ortega, 2009; Skehan, 2009).

Crucially, the pattern is theoretically meaningful under the "journey/touchpoints" lens: touchpoints are not interchangeable tasks but structured episodes that elicit different resource-integration behaviors, and therefore different c-CAF signatures (Lemon & Verhoef, 2016). In practical terms, short functional tasks appeared to foreground efficiency and local control (higher WPM; relatively favorable accuracy in particular touchpoints), whereas extended essays foregrounded discourse building and connective texture (higher cohesion and complexity). This supports the study's broader claim that value realization in EMI writing is contextual and episode-dependent, shaped by the design of touchpoints across the semester journey (Zomerdijk & Voss, 2010).

RQ3 clarifies why these touchpoint-dependent profiles emerge by identifying which interlanguage constraints persist and which are sensitive to genre demands. High-frequency error families—particularly articles/determiners, lexical choice/word form, and tense/aspect—were persistent across tasks, indicating stable vulnerability points in the developing system. Such persistence is consistent with interlanguage accounts that certain subsystems remain chronically unstable even as learners expand their expressive repertoire (Selinker, 1972). At the same time, the error ecology was systematically genre-sensitive: essays were associated with reduced article/determiner and preposition errors but increased lexical choice/word form errors, tense/aspect errors, and sentence boundary/punctuation problems (Table 3). One plausible interpretation is that essays intensify demands on lexical precision, temporal framing, and clause linkage—domains where L2 writers often experience instability—while simultaneously providing more opportunity for monitoring some function-word choices. This dual pattern (persistence plus task-conditioned reweighting) is consistent with the claim that accuracy measures and error distributions are highly contingent on task design and coding choices, requiring careful interpretation in L2 writing research (Polio, 1997). Taken together, RQ3 suggests that the observed gains in cohesion and complexity in essays were accompanied by newly amplified constraints in lexical and discourse-boundary control, highlighting the mechanism-level cost of more sophisticated discourse production under extended genre demands.

The study contributes in three ways. First, it operationalizes "value-in-use" in an educational setting through behavioral-linguistic traces rather than relying solely on perceptions, showing how value creation can be empirically observed as change and variation in performance across touchpoints (Vargo & Lusch, 2004). Second, it introduces a journey-based framing to EMI writing development by treating course tasks as touchpoints that systematically shape outcome profiles, thereby extending customer journey logic into an assessment-rich learning service context (Lemon & Verhoef, 2016). Third, it specifies interlanguage as the learner-side capability mechanism that mediates touchpoint effects, demonstrating how persistent constraints can coexist with robust development in discourse and lexical resources. Conceptually, this closes the loop of the proposed framework by showing that touchpoint design predicts not only aggregate performance but also the distribution of constraints and gains across dimensions.

4.2 Implications: designing EMI writing as a service journey (instruction, feedback, assessment)

The findings imply that EMI writing curricula should be treated as service journeys with differentiated touchpoints rather than as a homogeneous sequence of "writing assignments." Three actionable implications follow.

Balance expansion and stabilization touchpoints. Because complexity/lexical diversity/cohesion rose more

consistently than accuracy, course journeys may benefit from deliberately interleaving "expansion" touchpoints (extended essays that foster discourse building) with "stabilization" touchpoints that target persistent error families (articles/determiners; tense/aspect; lexical choice) without sacrificing communicative purpose.

Differentiate feedback by touchpoint profile. The error results suggest that short tasks are efficient venues for providing feedback on function words and local accuracy, whereas essay touchpoints warrant targeted support for lexical precision, tense/aspect control, and sentence boundary management—issues that become more salient when students attempt longer argumentative chains (Table 3).

Interpret development multidimensionally. Gains in cohesion and lexical diversity during essay writing should be interpreted as meaningful development in discourse-functional competence rather than as mere "verbosity," while small declines in error-free production can reflect increased linguistic ambition and complexity under higher genre demands. This reinforces the need for assessment rubrics and feedback systems that reward communicative effectiveness and discourse control alongside local accuracy.

4.3 Limitations and future research

Several limitations should be considered. First, the study was conducted in a single EMI writing course; generalizability to other EMI programs depends on differences in students' entry proficiency, disciplinary writing norms, and language policy environments (Macaro et al., 2018). Second, while c-CAF indicators provide a scalable, theory-aligned operationalisation, accuracy and cohesion measures can be sensitive to operational definitions. Future work should triangulate computational indices with human ratings and more fine-grained discourse coding to strengthen interpretive validity (Polio, 1997; Graesser et al., 2004). Third, to fully leverage the consumer/service positioning, future research should incorporate experience-side measures (e.g., perceived value, task difficulty, effort, and autonomy support) and test whether these perceptions mediate the link between touchpoint design and performance outcomes, as implied by customer journey theorizing (Lemon & Verhoef, 2016). Finally, future modeling could explore heterogeneity via cross-level interactions (e.g., baseline proficiency × touchpoint type) and examine whether persistent error families represent stable interlanguage constraints or strategic shifts under time pressure across genres (Selinker, 1972; Skehan, 2009).

5. CONCLUSION

This study examined EMI writing development as value-in-use enacted across a semester-long journey of genre touchpoints, with learners' interlanguage functioning as the mechanism through which course resources were converted into observable writing performance. Using repeated, timed writing tasks (20-minute functional genres and 45-minute essay genres) and a multidimensional outcome space (accuracy, fluency, syntactic complexity, lexical diversity, and cohesion), the findings provide a coherent account of how performance evolves and varies under authentic classroom constraints.

Three conclusions follow. First, across the semester, students demonstrated clear development in dimensions associated with linguistic elaboration and discourse organization: syntactic complexity, lexical diversity, and cohesion increased reliably over time, indicating that learners progressively expanded their capacity to produce longer, more lexically varied, and more textually connected academic writing. In contrast, accuracy did not improve monotonically, and fluency (writing rate) did not show a uniform upward trend, underscoring that interlanguage development may be most visible in expansion of expressive resources before stable reductions in error become apparent.

Second, writing outcomes were strongly touchpoint- and timing-dependent. Compared with 20-minute tasks, 45-minute essay tasks elicited substantially greater text production and higher complexity, lexical diversity, and cohesion, while writing rate decreased and error-free production declined slightly. These results support the interpretation that touchpoints are not interchangeable assessments: they are structured episodes that elicit different resource-allocation patterns and therefore distinct c-CAF profiles, with extended essay conditions amplifying discourse-level demands and lexical risk.

Third, interlanguage analysis revealed a dual pattern of persistence and genre sensitivity. Articles/determiners, lexical choice/word form, and tense/aspect errors remained high-frequency constraints across the corpus, indicating persistent pressure points in the developing system. At the same time, essay touchpoints shifted the error ecology: function-word errors (articles, prepositions) decreased, whereas lexical choice, tense/aspect, and sentence-boundary problems increased. This suggests that extended, argumentative writing magnifies lexico-grammatical and discourse-boundary pressures, even while some local accuracy features may stabilize.

The study contributes by operationalizing "value-in-use" in EMI writing through behavioral-linguistic traces at the touchpoint level rather than relying solely on perceptions. Practically, the findings imply that EMI writing curricula should be treated as service journeys that intentionally balance touchpoints for (a) discourse expansion and genre mastery and (b) stabilization of persistent interlanguage constraints. Future work should extend this approach by integrating experience measures (e.g., perceived value, effort, and task difficulty), modeling individual-difference moderators, and triangulating computational indices with human ratings to further strengthen validity and interpretability across EMI contexts...

REFERENCES
1. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects

for subjects and items. Journal of Memory and Language, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005 (ScienceDirect)

2. Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. Frontiers in Psychology, 4, 328. https://doi.org/10.3389/fpsyg.2013.00328 (Frontiers)

3. Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. Journal of Second Language Writing, 32, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

4. Dafouz, E., & Smit, U. (2020). ROAD-MAPPING English medium education in the internationalised university. Palgrave Macmillan. DOI: 10.1007/978-3-030-23463-8

5. Dearden, J. (2014). English as a medium of instruction – a growing global phenomenon (Phase 1). British Council.

6. Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. Behavior Research Methods, Instruments, & Computers, 36(2), 193–202.DOI: 10.3758/BF03195564

7. Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. Longman. https://doi.org/10.4324/9781315836010

8. Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA. John Benjamins. https://doi.org/10.1075/lllt.32

9. Hyland, K. (2004). Genre and second language writing. University of Michigan Press.

10. Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59(4), 434–446.

11. Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. Journal of Marketing, 80(6), 69–96.

12. Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. Language Teaching, 51(1), 36–76. https://doi.org/10.1017/S0261444817000350

13. Molesworth, M., Nixon, E., & Scullion, R. (2009). Having, being and higher education: The marketisation of the university and the transformation of the student into consumer. Teaching in Higher Education, 14(3), 277–287. https://doi.org/10.1080/13562510902898841

14. 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

15. Patrício, L., Fisk, R. P., & Falcão e Cunha, J. (2008). Designing multi-interface service experiences: The service experience blueprint. Journal of Service Research, 10(4), 318–334. https://doi.org/10.1177/1094670508314264.

16. Payne, A. F., Storbacka, K., & Frow, P. (2008). Managing the co-creation of value. Journal of the Academy of Marketing Science, 36(1), 83–96. https://doi.org/10.1007/s11747-007-0070-0

17. Polio, C.G. (1997), Measures of Linguistic Accuracy in Second Language Writing Research. Language Learning, 47: 101-143. https://doi.org/10.1111/0023-8333.31997003

18. Rose, H., Sahan, K., Wei, M., Aizawa, I., Zhou, S., & Shepard, C. (2026). A systematic review of English medium instruction in higher education: An update of Macaro et al. (2018). System, 136, 103892.https://doi.org/10.1016/j.system.2025.103892

19. Selinker, L. (1972). Interlanguage. IRAL: International Review of Applied Linguistics in Language Teaching, 10(1–4), 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209

20. Swales, J. M. (1990). Genre analysis: English in academic and research settings. Cambridge University Press.

21. Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. Journal of Marketing, 68(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036

22. Võ Đoàn T., & Ngô, H. H. N. (2025). Khung lí thuyết ROAD-MAPPING cho các nghiên cứu về các chương trình giảng dạy bằng tiếng Anh trong giáo dục đại học Việt Nam. Tạp Chí Giáo dục, 25(12), 47–52. https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/3506

23. Yao, C. W., Collins, C., Bush, T., Briscoe, K. L., & Dang, N. L. T. (2022). English as a "double barrier": English medium instruction and student learning at Vietnamese transnational universities. Higher Education Research & Development, 41(4), 1372–1386. https://doi.org/10.1080/07294360.2021.1896485

24. Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means–end model and synthesis of evidence. Journal of Marketing, 52(3), 2–22. https://doi.org/10.1177/002224298805200302

25. Zomerdijk, L. G., & Voss, C. A. (2010). Service design for experience-centric services. Journal of Service Research, 13(1), 67–82. https://doi.org/10.1177/1094670509351960

.