Individual Differences in Language Education: An International Journal

EUROKD

2025 (3), 14–29
https://doi.org/10.32038/idle.2025.03.02

Individual Differences in Task-Based Language Teaching: A Critical Synthesis for AI-Mediated Learning Design

Yanzhen Tan
Southeast University Chengxian College, China
Yawen Han
School of Foreign Languages, Southeast University, China
Zhisheng (Edward) Wen*
Division of Languages and Communication, The Hong Kong Polytechnic University - CPCE, Hong Kong, China

Correspondence
Email: edward.wen@cpce-polyu.edu.hk

Abstract

The integration of artificial intelligence (AI) into task-based language teaching (TBLT) promises personalized learning at scale. Yet without grounding in the rich tradition of individual differences (IDs) research, AI risks reducing learners to simplified variables and substituting algorithmic accommodation for genuine cognitive engagement. This paper builds a bridge between established ID-TBLT research and emerging AI design. We critically synthesize how cognitive IDs (working memory, attentional control, language aptitude) and affective IDs (motivation, anxiety, enjoyment, boredom, and flow) interact with task complexity to shape L2 performance—acknowledging both established findings and persistent debates. From this synthesis, we articulate design challenges for AI-mediated TBLT, distinguishing what is technically feasible from what requires further research, and acknowledging inherent tensions (scaffolding vs. substitution, personalization vs. equity, detection vs. privacy, flow vs. instructional efficiency). We then extend this framework through Bui's (2026) longitudinal study of learner perceptions in AI-assisted speaking tasks, using empirical findings on learner adaptation and "prompt literacy" to illustrate and refine our design challenges. The result is a dual contribution: a systematic translation of ID-TBLT research into AI design challenges, and a refined, six-frontier research agenda for developing intelligent, adaptive TBLT systems that are cognitively grounded, affectively attuned, and ethically responsible.

ARTICLE HISTORY
Received: 01 August 2025
Revised: 12 November 2025
Accepted: 01 December 2025

KEYWORDS
Individual Differences, TBLT, Task Complexity, Working Memory, AI in Language Education

How to cite this article (APA 7th Edition):
Tan, Y., Han, Y., & Wen, Z. E. (2025). Individual differences in task-based language teaching: A critical synthesis for AI-mediated learning design. *Individual Differences in Language Education: An International Journal, 3*, 14–29. https://doi.org/10.32038/idle.2025.03.02

[1]Introduction: Building a Bridge between Old and New

Task-based language teaching (TBLT) has, for decades, recognized that learners are not a homogeneous group. The field has accumulated a rich body of knowledge about how individual differences (IDs)—cognitive and affective—interact with task design to shape learning outcomes (Li et al., 2022; Wen et al., 2023). We know that working memory capacity moderates how learners manage cognitive load under complexity (Wen, 2016); that attentional control determines whether learners trade off accuracy for fluency or strategically allocate focus (Skehan, 2015); that language aptitude profiles predict responsiveness to different instructional conditions (Michaud & Ammar, 2023); that motivation ebbs and flows with task challenge (Kormos & Préfontaine, 2017); that anxiety can transform productive struggle into debilitating overload (Mora et al., 2024); and, more recently, that enjoyment, boredom, and flow—the optimal state of deep engagement—critically shape task performance (Dewaele & MacIntyre, 2014; Li et al., 2026).

This is the "old"—a sophisticated, empirically-grounded understanding of learner variation that the field has built over decades. Yet this knowledge remains contested in places, with ongoing debates about the magnitude of ID effects, the independence of constructs, and the generalizability of findings across contexts.

Now, the "new" has arrived. Generative AI and large language models are being rapidly integrated into language education, promising hyper-personalized learning environments that adapt to individual learners in real time. Yet this integration proceeds largely without reference to the ID research tradition. As the special issue editorial warns (Wen et al., 2025), an AI built without a cognitive model is shallow; an AI built on a *flawed* model of IDs could be actively harmful—entrenching inequities, creating dependencies, and eroding the productive struggle essential for learning.

This paper builds a bridge between old and new. By "old" we refer to the established body of ID-TBLT research synthesized in Sections 2 and 3; by "new" we refer to the emerging design challenges for AI-mediated learning articulated in Sections 4–6. We do not simply summarize ID-TBLT research, nor do we simply celebrate AI's potential. Instead, we ask: *What does decades of ID-TBLT research tell us about how AI should be designed?* And conversely: *What emerging evidence about AI-mediated learning tells us about how ID-TBLT research must evolve?*

[1] This paper is part of a special issue (2025, 3) entitled: Individual Differences in the AI and Digital Era: The Impact of ChatGPT and Beyond (edited by Zhisheng (Edward) Wen, Richard Sparks, and Hassan Mohebbi).

Our contribution is threefold. First, we provide a critical synthesis of key cognitive and affective IDs in TBLT, engaging with both established findings and persistent debates, and articulating what this research implies for AI design—distinguishing what is technically feasible from what requires further investigation. Second, we translate these implications into design challenges, acknowledging inherent tensions and trade-offs. Third, we extend this framework through Bui's (2026) longitudinal study of learner perceptions in AI-assisted speaking tasks, using empirical findings to illustrate and refine our design challenges. The result is a bridge that enables the field to move forward with both theoretical sophistication and empirical grounding.

A Critical Synthesis of Cognitive IDs in TBLT

We begin with cognitive IDs—the processing capacities that determine how learners manage the cognitive demands of tasks. For each ID, we synthesize key findings, acknowledge limitations and debates, and articulate implications for AI design as *design challenges* requiring further investigation.

Working Memory: Managing Cognitive Load Dynamically

Working memory (WM)—the limited-capacity system for simultaneous storage and processing (Baddeley, 2007)—is a critical moderator of task performance. Meta-analytic evidence confirms a modest but robust correlation ($r \approx .30$) between WM and L2 proficiency (Linck et al., 2014; Wen, 2016). This means WM explains approximately 9% of variance in outcomes—significant but not deterministic.

Key *findings and debates*:

- *WM effects are differentiated, not unitary*. Tasks that increase concurrent memory load (e.g., maintaining multiple narrative elements) interact with WM's storage components; tasks that increase reasoning demands load the central executive (Kormos & Trebits, 2011). However, the precise mapping between task features and WM components remains contested.
- *WM effects are contingent on task design*. Some studies show WM advantages emerge primarily under high complexity (Awwad & Tavakoli, 2022); others show advantages under simpler conditions (Kormos & Trebits, 2011). Null findings also exist (Cho, 2018), suggesting that not all complexity manipulations engage WM.
- *WM may be a correlate, not a cause*. The direction of causality between WM and L2 proficiency remains debated. WM may facilitate learning, but successful learning may also expand WM capacity for language tasks (Schwieter & Wen, 2022).

Attentional Control: Managing the Form-Meaning Trade-off

Attentional control—the goal-directed regulation of cognitive focus (Draheim et al., 2022)—is central to TBLT's theoretical debates. Two influential hypotheses appear contradictory. Skehan's (2015) Limited Attentional Capacity hypothesis predicts trade-offs (e.g., between meaning and form) as task demands increase. Robinson's (2011) Cognition Hypothesis posits that increasing complexity can selectively direct attention to form.

Key Findings and Debates:

- *Both hypotheses capture valid phenomena.* Trade-offs manifest in *online performance* under high load; attentional enhancement describes *developmental outcomes* when complexity successfully directs attention (Skehan, 2015; Robinson, 2011). However, these are post-hoc reconciliations; direct tests remain limited.
- *Attentional allocation is measurable through process data.* Eye-tracking, keystroke logging, and speech pauses can reveal where attention is directed moment-by-moment (Xing, 2023). But these measures remain largely laboratory-based.
- *Learners differ in attentional control capacity.* Individual differences in executive function affect ability to manage form-meaning competition (Miyake & Friedman, 2012).

Language Aptitude: Matching Instruction to Cognitive Profile

Language aptitude has evolved from Carroll's (1990) classic model—phonetic coding, grammatical sensitivity, inductive learning, rote memory—to include *implicit aptitude* for unconscious pattern detection (Granena, 2019; Suzuki, 2021). This evolution reflects a growing recognition that aptitude is multi-componential and context-dependent.

Key findings and debates:

- *Aptitude effects are conditional, not absolute.* The effectiveness of a given task design depends on alignment with the learner's aptitude profile (Kourtali & Révész, 2020).
- *Michaud and Ammar (2023) demonstrated a three-way interaction:* analytic aptitude predicted learning for lower-proficiency learners receiving explicit instruction; implicit aptitude predicted learning for higher-proficiency learners in implicit conditions. This suggests aptitude effects are mediated by proficiency and instructional context.
- *Aptitude measurement remains contested.* The LLAMA battery, while widely used, has known limitations; LLAMA D's construct validity as a measure of implicit aptitude is debated (Suzuki, 2021). The MLAT remains the gold standard but is not language-neutral.

Table 1 presents a consolidated overview of design challenges for cognitive IDs.

Table 1

Design Challenges for Cognitive IDs in AI-Mediated TBLT

ID Construct	Key Findings	Design Challenge	Feasibility
Working Memory	WM effects are differentiated (storage vs. processing); contingent on task design; may be correlate, not cause	Distinguish storage from processing demands; infer WM load in real time; provide compensatory scaffolding that fades over time	Research prototype; NLP-based inference feasible; validation needed
Attentional Control	Trade-offs in performance; enhancement in development; allocation measurable via process data	Monitor attentional allocation via multimodal data; detect form-meaning trade-offs; deliver just-in-time prompts without disrupting flow	Eye-tracking not scalable; other indicators feasible; research area for detection
Language Aptitude	Effects are conditional; three-way interaction with proficiency and instruction; measurement contested	Profile multiple aptitude dimensions (explicit/implicit); branch learning pathways; adapt feedback to aptitude profile	Stealth assessment promising but unvalidated; efficacy untested

A Critical Synthesis of Affective IDs in TBLT

Affective IDs determine not only *what* learners can do but *whether* they engage at all. For TBLT, this means tasks must be designed not only for cognitive optimization but also for motivational sustainability and affective safety. Recent research has expanded the affective landscape beyond motivation and anxiety to include **enjoyment**, **boredom**, and—most integratively—**flow**, the optimal state where deep engagement and positive affect converge (Li et al., 2026).

Motivation: Sustaining Engagement through Task Design

Motivation has evolved from trait-oriented models (Gardner, 1985) to dynamic, person-in-context frameworks. Dörnyei's (2005) L2 Motivational Self System emphasizes the *ideal L2 self*—a positive vision of proficiency—alongside the immediate *L2 learning experience.*

Key findings and debates:

- *Motivation is situated and fluctuating, not static.* Learners' engagement rises and falls within tasks in response to perceived challenge, autonomy, and relevance (Kormos & Préfontaine, 2017). However, most research relies on self-report, which captures reflections after the fact, not real-time dynamics.
- *Task complexity moderates motivational effects.* Optimal challenge energizes engagement; overwhelming complexity triggers disengagement. But "optimal" is highly individual and context-dependent.
- *The L2 Motivational Self System has been critiqued* for its reliance on self-report and its cultural specificity. Cross-cultural validation remains limited.

Anxiety: Managing the Affective Filter
Language anxiety—tension and apprehension specific to L2 communication (Horwitz et al., 1986)—functions variably as mediator, moderator, and outcome in TBLT.

Key findings and debates:
- *Increased task complexity can indirectly impair performance* by elevating anxiety, which consumes cognitive resources (Mora et al., 2024). However, the direction of causality is contested: does complexity cause anxiety, or do anxious learners perceive tasks as more complex?
- *Anxiety moderates the complexity-performance relationship:* low-anxiety learners may benefit from increased complexity; high-anxiety learners may experience breakdown (Kim & Tracy-Ventura, 2011).
- *The anxiety-complexity link is nonlinear*. Exceeding an individual's anxiety threshold triggers sharp performance declines. But thresholds are individual and dynamic.

Enjoyment, Boredom, and Flow: The Affective-Cognitive Interface
Recent research has expanded the affective landscape to include **enjoyment** and **boredom** as distinct, consequential emotions (Dewaele & MacIntyre, 2014; Li, 2021; Pawlak et al., 2020). More recently, **flow**—the optimal state characterized by deep absorption, loss of self-consciousness, and intrinsic enjoyment—has emerged as a critical construct that bridges affective and cognitive dimensions of task engagement (Li et al., 2026; Csikszentmihalyi, 1990).

Key findings and debates:
- *Enjoyment is not merely the absence of anxiety*. Enjoyment and anxiety are weakly correlated or independent, with distinct predictors: enjoyment is associated with teacher support, task variety, and perceived competence; anxiety with social evaluation and task difficulty (Dewaele & MacIntyre, 2014; Li, 2021).
- *Boredom is a distinct and prevalent emotion*. L2 boredom is associated with inattention, reduced effort, and poorer learning outcomes (Pawlak et al., 2020). It arises from under-challenge, repetition, and perceived irrelevance—conditions that AI systems may inadvertently create.
- *Flow represents the optimal affective-cognitive state.* Grounded in Csikszentmihalyi's (1990) flow theory and Egbert's (2003) model of flow in language learning, flow is characterized by intense focus, balance between challenge and skill, clear goals, immediate feedback, and intrinsic enjoyment. In task-based contexts, flow predicts persistence, creativity, and performance (Li et al., 2026).
- *Flow is distinct from both boredom and anxiety*. Flow occurs when challenge matches skill; boredom when challenge is too low; anxiety when challenge is too high. This three-channel model has direct implications for task design.
- *Flow mediates the effects of task conditions on performance.* Li et al. (2026)

demonstrated that perceived task control predicts writing performance both directly and *indirectly through flow*. This establishes flow as a key mechanism linking task design to outcomes.

Debates and limitations:

- The relationship between enjoyment, boredom, and flow is complex. Flow is not simply the sum of high enjoyment and low boredom; it is a distinct state with unique antecedents.
- Most research relies on self-report questionnaires, which capture global experiences rather than moment-to-moment flow fluctuations.
- The conditions that promote flow may vary across individuals and tasks.

Unlike enjoyment and boredom, which can function as both momentary states and stable dispositions, flow is conceptualized as a momentary state arising from the alignment of task challenge and learner skill (Csikszentmihalyi, 1990). This distinction has implications for AI design: systems must detect flow in real time, not assume trait-level dispositions. Table 2 presents a consolidated overview of design challenges for affective IDs.

Table 2

Design Challenges for Affective IDs in AI-Mediated TBLT

ID Construct	Key Findings	Design Challenge	Feasibility
Motivation	Situated and fluctuating; task complexity moderates effects; autonomy supports engagement	Track motivational indicators (time-on-task, help-seeking); personalize challenge; support autonomy through choice	Behavioral proxies feasible; validation needed; efficacy untested
Anxiety	Mediates complexity effects; nonlinear; thresholds vary individually	Detect anxiety in real time (with privacy safeguards); implement graduated exposure; provide just-in-time safety interventions	Privacy concerns; accuracy contested; feasible in principle
Enjoyment, Boredom, Flow	Flow distinct from boredom/anxiety; flow mediates task effects; flow is momentary state, not trait	Detect flow, enjoyment, boredom; calibrate challenge to skill; provide clear goals and immediate feedback; maintain challenge-skill balance over time	Behavioral indicators feasible; validation needed; timing critical

Bridging Old and New: Design Challenges for AI-Mediated TBLT

Synthesizing the implications above, we articulate five core design challenges for AI-mediated TBLT. These are not ready-to-implement principles but research frontiers requiring systematic investigation. Table 3 summarizes these challenges with their key research questions.

Table 3

Summary of Design Challenges and Key Research Questions

Design Challenge	Foundation	Key Research Question
1. Differentiated Cognitive Load Management	WM effects are differentiated; tasks load storage vs. processing differently	Can we develop multimodal indicators that reliably distinguish storage from processing load?
2. Dynamic Attentional Scaffolding	Attention allocation is measurable; form-meaning trade-offs occur	What multimodal data streams serve as reliable indicators of attentional allocation, and how should prompts be timed?
3. Aptitude-Instruction Alignment	Aptitude profiles moderate responsiveness to instructional conditions	Can aptitude profiles be reliably inferred from task interaction data, and does matched instruction produce better outcomes?
4. Motivation-Sensitive Sequencing	Motivation is dynamic and responsive to task challenge	What behavioral indicators reliably predict waning engagement, and does adaptive sequencing maintain motivation?
5. Affect-Aware Safety Systems	Anxiety is nonlinear; flow requires challenge-skill balance	Can we develop affect detection that is accurate, privacy-preserving, and culturally sensitive, and does flow-sensitive sequencing enhance outcomes?

Acknowledging Tensions and Trade-offs

These design challenges interact in ways that create inherent tensions. Scaffolding vs. substitution: support that reduces cognitive load may also reduce productive struggle. Systems must fade support over time, and evaluation must measure independence, not just immediate performance. Personalization vs. equity: ID-based personalization may create differential learning opportunities. Personalization must aim to close gaps, not widen them. Detection vs. privacy: affect detection requires sensitive data. Privacy must be a first-order design constraint, with informed consent and data minimization. Flow vs. instructional efficiency: flow-producing tasks may not align with learning objectives; instructionally efficient tasks (maximizing learning outcomes per unit time) may not produce flow. Design must balance engagement with learning goals, evaluating both learning outcomes and affective-cognitive states. These tensions are not solvable by technical means alone; they require ongoing ethical reflection, stakeholder engagement, and empirical investigation.

Extending the Bridge: Bui's (2026) Empirical Illustration

The design challenges above derive from traditional ID-TBLT research conducted largely *without* AI. But AI is now here, and learners are using it. What does emerging evidence about AI-mediated learning tell us about these challenges?

Bui's (2026) seven-week longitudinal study of L2 learners interacting with AI chatbots versus human interlocutors provides a valuable empirical illustration. Table 4 summarizes Bui's findings and their implications for our design challenges.

Table 4
Bui's (2026) Findings: Phases, Illustrations, and Refinements

Phase	Key Characteristics	Illustrative Quotes
Phase 1: Enthusiastic Adoption	Output stimulation, ease of access, lower affective barriers	"I can just talk without worrying about being judged."
Phase 2: Lukewarm Engagement	Lack of social presence, speech prosody, responsive interaction	"It gets repetitive. The AI doesn't really respond to what I say."
Phase 3: Strategic, Purpose-Driven Use	Prompt literacy; AI for bounded purposes; humans for authentic communication	"I use AI for fluency practice. For real conversation, I prefer a person."

What Bui Illustrates	What Bui Adds (Refinements)
Affect-Aware Safety Systems (AI reduces affective barriers)	Social Presence: Transition learners from AI to human interaction
Motivation-Sensitive Sequencing (trajectory shows dynamic motivation)	Prompt Literacy: Systems should develop this strategic competence
Aptitude-Instruction Alignment (individual differences in strategic use)	Hybrid as Default: AI is one component of a hybrid system
	Flow Disruption: Maintain challenge-skill balance over time

Limitations: This is a single qualitative study with a specific AI tool. Findings are suggestive, not generalizable. The novelty effect (Clark & Mayer, 2016) may explain part of the trajectory. Additionally, the study documents *perceptions*, not *learning outcomes*.

A Refined Research Agenda: Six Provocations for the Field

Building on the synthesis above, we propose six research frontiers that move beyond generic calls for "more research" to offer provocations—specific, timely, and theoretically generative directions that emerge from the intersection of ID-TBLT research, the editorial's CLWM framework, and cutting-edge empirical work on AI-mediated learning.

Frontier 1: From Static Tests to Stealth Assessment—Can We Read IDs in Interaction?

For decades, we have assessed working memory with span tasks, aptitude with the MLAT, and anxiety with questionnaires administered before or after learning. These tools are valuable but fundamentally offline—they capture traits, not states; they measure capacity, not real-time deployment. AI offers the possibility of stealth assessment: inferring ID-relevant states from the behavioral traces learners leave as they interact with tasks.

The timely provocation here is not whether we *can* do this—preliminary work exists—but whether we *should*, and if so, under what conditions. Bui's (2026) learners developed prompt literacy spontaneously; their strategic adaptations were visible in how they used AI. Could we have detected their evolving competence through interaction patterns alone? More provocatively, could we have inferred their WM load from speech hesitancy, their anxiety from pause patterns, their flow from sustained focus?

The timely questions: Can we develop multimodal indicators (speech hesitancy, response latency, repetition errors, pause dynamics, keystroke patterns) that reliably distinguish storage overload from executive overload in real time? Can we infer implicit versus analytic aptitude from how learners respond to feedback—whether they correct errors after explicit rules or after implicit recasts? How do AI-inferred metrics correlate with established ID measures, and critically, do they predict learning outcomes with greater ecological validity?

Why this is trendy: Stealth assessment is moving from research prototype to practical possibility with advances in NLP and multimodal analytics. But the field risks reinventing wheels—developing indicators without theoretical grounding. Our provocation is to ground stealth assessment in the established ID constructs we already understand, using AI to make visible what has always been there.

Frontier 2: The Personalization Paradox—Does Adapting to IDs Actually Help?
The promise of AI is personalization. But personalization rests on a premise rarely tested: that adapting to individual differences produces better outcomes than well-designed one-size-fits-all instruction. This is the personalization paradox: the more we personalize, the less we know whether personalization itself matters.

The timely provocation is to test the null hypothesis. Do systems that adapt to WM load actually outperform non-adaptive systems—or do learners adapt to whatever system they encounter, developing strategies that compensate for its limitations (as Bui's learners did with AI)? Does aptitude-matched instruction produce better outcomes than mismatched instruction, or does exposure to varied instructional conditions build flexibility that matched instruction forecloses? Does graduated exposure sequencing—AI scaffolding before human interaction—reduce anxiety and improve outcomes, or does it delay the productive struggle that builds resilience?

Why this is trendy: Large language models are being deployed in education with minimal efficacy testing. The field needs rigorous randomized controlled trials that systematically test interaction effects: do ID-sensitive systems benefit some learners more than others? For whom does personalization work, and for whom does it fail? The provocation is to treat personalization as a hypothesis to be tested, not a solution to be implemented.

Frontier 3: Prompt Literacy—A New Individual Difference for the AI Era?
Bui's (2026) most striking finding was that learners developed prompt literacy—the strategic ability to deploy AI for specific purposes—spontaneously, without instruction. This competence varied across individuals. Some learners became sophisticated strategists, using AI for fluency practice and vocabulary acquisition while turning to human partners for authentic communication. Others remained passive consumers, frustrated by AI's limitations.

The timely provocation is to ask: Is prompt literacy a new individual difference? Does it correlate with established IDs—working memory, metacognitive awareness, analytic aptitude—or is it a distinct competence that the AI era demands? Can it be taught, and if so, what interventions accelerate its development? Most provocatively, does prompt literacy predict learning outcomes in AI-mediated instruction controlling for other IDs?

Why this is trendy: As AI becomes ubiquitous, the ability to use it strategically may matter as much as traditional cognitive abilities. The provocation is to treat prompt literacy not as a technical skill but as a cognitive competence—one that intersects with established ID constructs in ways we are only beginning to understand. Bui's learners showed us what it looks like to develop this competence; now we need to understand how to cultivate it.

Frontier 4: The Hybrid Imperative—Beyond AI-Only or Human-Only

The editorial argues that AI should complement, not replace, human interaction. Bui's (2026) learners confirmed this empirically: they consistently preferred human interlocutors for authentic communication, valuing what AI could not provide—social presence, speech prosody, responsive interaction. Yet we know almost nothing about **how** to sequence AI and human interaction optimally.

The timely provocation is to move beyond the false binary of AI-only versus human-only and ask: What does optimal hybrid pedagogy look like? Does AI serve best as preparation for human interaction (scaffolding fluency before authentic conversation), or as follow-up practice after human interaction (consolidating what was learned)? Does the optimal sequence vary by learning goal—AI for fluency, humans for pragmatic competence—or by ID profile—longer AI scaffolding for high-anxiety learners, faster transition for high-WM learners?

Why this is trendy: The AI-in-education discourse is polarized between techno-utopians who envision AI replacing teachers and skeptics who reject AI entirely. The hybrid imperative offers a third path—one that takes seriously both AI's affordances and its limitations. The provocation is to design and rigorously test hybrid pedagogies, treating AI and human interaction as complementary tools to be sequenced, not competing alternatives to be chosen.

Frontier 5: Algorithmic Gatekeeping—Ethics as Design Constraint, Not Afterthought

The editorial warns that an AI built without a cognitive model is shallow; an AI built on a flawed model of IDs could be actively harmful. This is not abstract caution. AI systems deployed in education already exhibit bias: they may reinforce stereotypes, disadvantage non-dominant language varieties, and make decisions that teachers cannot understand or override.

The timely provocation is to treat ethics not as an afterthought but as a design constraint from the outset. This means asking: Do AI systems exhibit bias across proficiency levels, language backgrounds, gender, and demographics? What frameworks for data privacy and informed consent are appropriate for multimodal analytics that capture facial expression, voice stress, and eye-tracking? How can AI systems be designed for algorithmic transparency—so teachers and learners can understand *why* a system made a particular adaptation and override it when necessary? Most provocatively, does ID-based personalization close or widen achievement gaps? Under what conditions does personalization promote equity, and under what conditions does it exacerbate existing inequities?

Why this is trendy: AI ethics is no longer niche; it is central to public debate. But educational AI ethics has lagged behind, often treated as a matter of compliance rather than critical inquiry. The provocation is to center ethics in the research agenda from the beginning—auditing systems for bias, designing for transparency, and treating equity as a primary outcome, not a secondary consideration.

Frontier 6: Flow, Boredom, and the Affective-Cognitive Interface

Li et al.'s (2026) demonstration that flow mediates the effects of task design on performance is a watershed finding for TBLT. It establishes that how learners *feel* during tasks is not merely a byproduct but a mechanism—a pathway through which task conditions shape outcomes. Yet flow remains understudied in AI-mediated learning, and its relationship to cognitive load management (the editorial's CLWM framework) is unexplored.

The timely provocation is to investigate the affective-cognitive interface directly. Can AI systems detect flow, enjoyment, and boredom through behavioral indicators—elaboration, initiation, sustained focus, interaction patterns—combined with real-time self-report triggers? Do task features—complexity, variety, autonomy, goal clarity, feedback immediacy, social presence—predict these states, and how does this vary by ID? Does optimizing for flow enhance learning outcomes, or does it conflict with instructional efficiency?

Most provocatively: What is the relationship between flow and productive struggle? The editorial insists that learning requires struggle—cognitive effort, difficulty, the possibility of failure. Flow is often characterized as effortless absorption. Is flow therefore incompatible with productive struggle, or does it represent its optimal form—challenging enough to engage but not so overwhelming as to frustrate? Answering this question would directly extend the editorial's CLWM framework, connecting cognitive load management to the affective experience of learning.

Why this is trendy: The affective turn in SLA is well underway, but flow—the optimal state—has received less attention than anxiety or enjoyment. Li et al. (2026) study changes that, providing empirical evidence that flow matters for performance. The provocation is to take flow seriously as a design goal for AI-mediated TBLT, investigating how AI can maintain challenge-skill balance over time, preventing both the under-challenge that produces boredom and the over-challenge that produces anxiety.

Conclusion: Building the Bridge

We began with a premise: the integration of AI into TBLT must be grounded in the rich tradition of individual differences research. Without this grounding, AI risks reducing learners to simplified variables and substituting algorithmic accommodation for genuine cognitive engagement.

We have built a bridge between old and new. From decades of ID-TBLT research—acknowledging both established findings and persistent debates—we articulated five design challenges for AI-mediated TBLT: differentiated cognitive load management, dynamic attentional scaffolding, aptitude-instruction alignment, motivation-sensitive sequencing, and affect-aware safety systems. We acknowledged inherent tensions: scaffolding vs. substitution, personalization vs. equity, detection vs. privacy, flow vs. instructional efficiency.

We extended this framework by incorporating cutting-edge research on enjoyment, boredom, and flow—particularly Li et al.'s (2026) demonstration that flow mediates task design effects on performance. We then used Bui's (2026) longitudinal study as empirical illustration, showing how learners' trajectory from enthusiasm to strategic use validates our design challenges and reveals new dimensions: social presence, prompt literacy, hybrid imperative, and flow disruption.

The research agenda we propose—six provocations for the field—charts a path forward that is timely, refreshing, and theoretically generative. From stealth assessment to the personalization paradox, from prompt literacy as a new individual difference to the hybrid imperative, from algorithmic gatekeeping to the affective-cognitive interface of flow—these are not generic calls for more research but specific, contested, and exciting questions that emerge from the intersection of ID-TBLT research and AI-mediated learning.

The stakes are high. AI is being integrated into language education with unprecedented speed. Without theoretical grounding, this integration will be driven by technological capability rather than pedagogical wisdom. But with grounding—with a bridge between what we know and what we are building—we can create AI systems that are not merely efficient but genuinely empowering: systems that see, understand, and nurture the unique mind of every language learner, while respecting the irreplaceable value of human connection and the optimal experience of flow.

ORCID
https://orcid.org/0009-0007-2468-1102
https://orcid.org/0000-0001-8165-3991
https://orcid.org/0000-0001-9041-6920

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

The authors thank the editors and reviewers for their constructive feedback on earlier versions of this manuscript.

Funding

We received no fund for this research project.

CRediT Authorship Contribution Statement

Yanzhen Tan: Conceptualization, Writing - Original Draft, Writing - Review & Editing
Yawen Han: Conceptualization, Writing - Review & Editing
Zhisheng (Edward) Wen: Conceptualization, Writing - Original Draft, Writing - Review & Editing

Generative AI Use Disclosure Statement

Portions of the literature synthesis and the refinement of the research agenda were developed with the assistance of DeepSeek AI. All intellectual content, theoretical framing, critical analysis, and final editorial decisions remain the sole responsibility of the authors.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants

Not applicable due to the nature of the paper.

Competing Interests

The authors have no competing interests.

Data Availability

There is no data available because of the nature of the paper.

References

Awwad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching, 60*(2), 169–196. https://doi.org/10.1515/iral-2018-0378

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford University Press.
Bui, G. (2026). From Novelty to Strategy: The Trajectory of Learner Perceptions in Human and AI-Assisted L2 Speaking Tasks. *International Journal of TESOL Studies*, 8(1), 108-127. https://doi.org/10.58304/ijts.260301
Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude. In T. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–29). Prentice-Hall.
Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System, 72*, 85–98. https://doi.org/10.1016/j.system.2017.10.010
Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction* (4th ed.). Wiley.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5
Dörnyei, Z. (2005). *The psychology of the language learner*. Lawrence Erlbaum.
Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review, 29*(4), 1143–1197. https://doi.org/10.3758/s13423-021-02052-2
Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal, 87*(4), 499–518. https://doi.org/10.3138/cmlr.60.5.549
Gardner, R. C. (1985). *Social psychology and language learning*. Edward Arnold.
Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency. *Studies in Second Language Acquisition*, *41*, 313–336. https://doi.org/10.1017/S0272263118000256
Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x
Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity* (pp. 287–306). John Benjamins.
Kormos, J., & Préfontaine, Y. (2017). Affective factors influencing fluent performance. *Language Teaching Research*, *21*, 699–716. https://doi.org/10.1177/1362168816683562
Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity* (pp. 267–289). John Benjamins.
Kourtali, N., & Révész, A. (2020). The roles of recasts, task complexity, and aptitude in child second language development. *Language Learning*, *70*, 179–218.https://doi.org/10.1111/lang.12374
Li, C. (2021). A control-value theory approach to boredom in English classes among university students in China. *The Modern Language Journal, 105*(1), 317–334. https://doi.org/10.1111/modl.12693
Li, C., Dewaele, J.-M., & MacIntyre, P. (2026). Task flow in L2 writing: Antecedents and effects on task performance. *Studies in Second Language Learning and Teaching, 16*(1). https://doi.org/10.14746/ssllt.45224
Li, S., Hiver, P., & Papi, M. (Eds.). (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.
Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2
Michaud, G., & Ammar, A. (2023). The role of language aptitude and timing of form-focused instruction in TBLT. In Z. Wen, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (pp. 357–380). Cambridge University Press.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458
Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking anxiety and task complexity effects on second language speech. *International Journal of Applied Linguistics, 34*(1), 292–315. https://doi.org/10.1111/ijal.12494
Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System, 91*, Article 102259. https://doi.org/10.1016/j.system.2020.102259
Robinson, P. (2011). Second language task complexity. In P. Robinson (Ed.), *Second language task complexity* (pp. 3–38). John Benjamins.
Schwieter, J. W., & Wen, Z. (Eds.). (2022). *The Cambridge handbook of working memory and language*. Cambridge University Press.

Skehan, P. (2015). Limited attentional capacity and cognition. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 123–156). John Benjamins.
Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition, 43*(3), 663–676. https://doi.org/10.1017/S0272263120000704
Wen, Z. (2016). *Working memory and second language learning*. Multilingual Matters.
Wen, Z., Sparks, R. L., & Mohebbi, H. (2025, this issue). Toward cognitively grounded AI: Can a working memory interface bridge individual differences and system design? *Individual Differences in Language Education, 3*, x–xx.
Wen, Z., Sparks, R. L., Biedroń, A., & Teng, M. F. (2023). *Cognitive individual differences in second language acquisition*. De Gruyter. https://doi.org/10.1515/9781614514749
Xing, J. (2023). How do EFL learners process oral tasks with different complexity: An exploratory study. *Frontiers in Psychology*, *14*, 1–8. https://doi.org/10.3389/fpsyg.2023.1241964