

Stage-Based Dynamic Assessment for Personalized Grammar Instruction: A Case Study with EFL University Students

Kseniya V. Yatskina^{1*}

¹ RUDN University, Russian Federation

*Corresponding author's email: yatskina-kv@rudn.ru

 <https://orcid.org/0009-0009-6399-5985>

 <https://doi.org/10.54855/ictcp.26111>

® Copyright (c) 2026 Kseniya V. Yatskina

Received: 15/05/2026

Revision: 01/08/2026

Accepted: 06/08/2026

Online: 10/08/2026

ABSTRACT

Keywords: Zone of Proximal Development (ZPD), dynamic assessment, stage-based dynamic assessment, EFL, AI-powered tool, SFS-LPS model

The inherent heterogeneity of university EFL classrooms poses a significant challenge: students at the same CEFR level exhibit disparate gaps in their mastery of grammatical structures, yet traditional instruction remains one-size-fits-all. Current AI-powered tools, while effective at personalizing content sequencing, fail to diagnose the qualitative nature of individual difficulties and assess a learner's responsiveness to mediation. This study addresses the need for digital tools grounded in Vygotsky's Zone of Proximal Development (ZPD) and Bernstein's skill formation stages (SFS) to capture both learners' current mastery and their learning potential. A case study involving university EFL students was conducted utilizing a digital, stage-based dynamic assessment tool. By incorporating a four-level mediation system, the tool assessed each student's specific grammar mastery level (SFS), calculated their Learning Potential Score (LPS), generated individual learning profiles and assigned targeted tasks. This process operated within an iterative assessment-feedback-correction cycle. The findings revealed distinct learning paths and continuous nature of the cycle enabled ongoing monitoring and timely instructional adjustments. The theoretically grounded framework thus offers a scalable approach to personalizing instruction in EFL classrooms.

Introduction

University EFL classrooms are marked by a high degree of learner heterogeneity. Even when students are placed at the same CEFR level, their underlying grammatical competence often reveals a patchwork of unevenly developed skills. Conventional grammar instruction, however, tends to follow a uniform syllabus, treating all learners within the same proficiency band as if their developmental needs were identical. This mismatch between instructional uniformity and learner diversity limits the efficiency of classroom teaching and leaves many students without the targeted support they need.

The recent proliferation of AI-driven adaptive platforms has partially addressed this gap by personalizing the sequencing and difficulty of language tasks (Vo Thi Quynh Nga & Tran Kieu My An, 2025). These tools predominantly rely on accuracy-based metrics: they detect mistakes and adjust the task level accordingly. What they generally fail to capture is the qualitative nature of learners' difficulties and, crucially, their modifiability – the degree to which they can benefit from instructional assistance. As a result, current systems operate within a static assessment model that overlooks a dimension which has been central to Vygotsky's (1935) zone of proximal development (ZPD) and dynamic assessment (DA) methodology: the distance between independent performance and performance under mediation.

This limitation has motivated the development of a Stage-Based Computerized Dynamic Assessment (C-DA) platform that integrates two complementary constructs: the Skill Formation Stage (SFS) framework and the Learning Potential Score (LPS). The SFS framework, derived from Bernstein's (1947) theory of stage-by-stage skill development and adapted for language learning, describes the developmental trajectory of grammatical skill from elementary reproduction to creative application. The LPS provides a quantitative measure of learner responsiveness to graduated mediation, capturing the effort required to move from an incorrect initial response to a correct one following assistance.

The platform operationalizes these constructs through a four-level mediation system that assesses each student's current SFS, calculates their LPS, generates individual learning profiles, and assigns targeted tasks. This process operates within iterative assessment-feedback-correction cycles. A prior pilot study has empirically validated the platform's effectiveness in fostering grammatical development, demonstrating its superiority over both traditional instruction and adaptive AI-powered tools (Akhrenova & Yatskina, 2026). The current paper shifts the analytical lens to the diagnostic core of the Stage-Based C-DA platform: how the SFS-based classification relates to learners' mediation needs. To establish the validity of this classification, the analysis pursues two complementary aims. First, it examines how learners with different dominant SFS levels (A-D) differ in the quantity and quality of mediation required to complete tasks on their dominant level. Second, it explores the framework's developmental potential by tracing the evolution of SFS and LPS across three C-DA cycles through two contrasting cases.

Taken together, these analyses demonstrate the framework's capacity to generate individual learning grammar profiles and differentiate learners' mediation needs, while also offering preliminary insights into how these profiles may evolve over time. The findings are presented as descriptive patterns observed within this pilot case study sample, with the intention of informing future large-scale validation research.

Literature review

Grammar Teaching and Assessment

Although grammar is recognized as a fundamental component of language systems (Yarkofoji, 2019), governing structural organization and the relationships among linguistic elements, grammar instruction has always been one of the most controversial and least understood aspects of language teaching (Thornbury, 1999). In the field of language pedagogy, approaches to grammar instruction have evolved from grammar-translation and audio-lingual methods to communicative language teaching (CLT) (Nasai, 2025).

Recent studies, notably a survey of 231 teachers from 18 countries conducted by Burns (2009),

document a broad professional consensus: the vast majority of educators reject isolated grammar teaching and consider its integration into a meaningful communicative context to be the most effective strategy. This global trend, however, receives a more nuanced interpretation in the work of Schurz and Coumel (2023), who, drawing on a sample of 615 teachers from three European countries (Sweden, Austria and France), demonstrate that the specific forms of implementing such integration vary considerably depending on the national context and educational level.

An ambiguous picture emerges in the assessment of grammatical competence (Purpura, 2004). Purpura identifies several historically evolved approaches to grammar assessment – from the rule recitation and translation, to the application of rules in language exercises, and, finally, to the communicative use of grammatical structures in real-world tasks. The key difficulty, according to the author, is that there is still no agreement on what exactly constitutes grammatical knowledge, nor on which types of tasks allow it to be measured reliably and validly.

Han (2025) notes that nowadays formative assessment of grammar skills generally plays a secondary role: grammar assessment is predominantly summative, reduced to recording a final outcome, while ignoring the tasks of continuous diagnosis and instructional correction. This persistent neglect of the learning process represents a critical gap in current assessment practices. This gap can be effectively bridged by dynamic assessment (DA) methodology, which integrates diagnosis and instruction into a single process and provides feedback that is sensitive to the learner's zone of proximal development (ZPD). Furthermore, the computerized implementation of DA (C-DA) addresses the scalability and standardization challenges inherent in traditional DA, making it a practical and powerful solution for wider educational application.

Computerized Dynamic Assessment (C-DA)

According to Vygotsky's (1935) classic definition, the zone of proximal development (ZPD) is the distance between the actual developmental level, as determined by independent problem solving, and the level of potential development, as determined through problem solving under adult guidance or in collaboration with more capable peers. In language pedagogy, this concept has been operationalized through the model of Dynamic Assessment (DA) (Lantolf & Poehner, 2010).

DA is a diagnostic research method whose central element is a specially organized mediation process. Unlike standardized testing, when difficulties or mistakes arise, the learner receives a sequential series of graduated prompts (from a recommendation to review the answer to a direct explanation of the relevant rule) (Aljaafreh & Lantolf, 1994). This approach makes it possible to quantify the amount and nature of required support: the fewer prompts needed to independently achieve the result, the closer the learner's actual developmental level is to their potential capabilities. Thus, the procedure provides an instrumental measurement of the parameters of the zone of proximal development (Poehner & Lantolf, 2021).

In the methodology of dynamic assessment, two main approaches are distinguished: *interventionist* and *interactionist*. The *interventionist approach* is characterized by a standardized system of pre-developed prompts, which ensures a high degree of controllability and reproducibility of the process. In contrast, the *interactionist approach* is based on flexible interaction between the teacher (mediator) and the learner, where mediation is shaped and adapted to the student's actual needs during the course of collaboration (Lantolf & Poehner, 2010).

Despite the proven diagnostic effectiveness, the wide application of dynamic assessment was

long constrained by its resource intensity. A natural response to this challenge was the DA digitalization. The development of computerized dynamic assessment (C-DA) automates the mediation process by fully or partially delegating the teacher's function to an algorithmic environment, in which a system of graduated prompts is implemented as a pre-scripted scenario. In this environment, a key metric becomes the Learning Potential Score (LPS), which combines the learner's initial independent response (Actual Score, AS) with the response attained after receiving the graduated prompts (Mediated Score, MS) to quantify the learner's developmental potential within the ZPD (Poehner, Zhang, & Lu, 2015).

Contemporary studies demonstrate the effectiveness of C-DA for diagnosing and developing various components of communicative competence (Qin & Zhang, 2019).

Thus, the study by Kamrood, Davoudi, Ghaniabadi, and Amirian (2021) demonstrated the effectiveness of C-DA for diagnosing and developing the listening skills of EFL students. In the experiment, 43 students completed 25 multiple-choice tasks with mediation in the form of graduated prompts. The results showed a statistically significant increase in scores after mediation, revealing that learners with identical initial scores responded differently to mediation, while detailed task performance profiles revealed individual strengths and weaknesses in specific listening constructs.

A broader C-DA project is presented in the study by Ebadi, Karimi, and Vakili (2023), who developed a web platform aimed at enhancing inferential listening skills while minimizing guessing effects. Ninety-four Iranian students completed 20 multiple-choice tasks with graduated mediation (four levels of prompts – from implicit to explicit) and a built-in guess-checker to verify the justification of correct answers. Quantitative analysis showed a statistically significant difference between actual and mediated scores, corroborating learners' sensitivity to help. Building on these works, Meng, Fu, and Wang (2024) proposed an attribute-based model of mediation (AMM) that enabled more precise diagnosis and development of specific components of listening ability.

The C-DA approach has also been successfully extended to another receptive skill – reading. Here, particular attention deserves the study by Siengyen and Wasanasomsithi (2024), which not only notes the benefits of mediation but also traces learners' progress in detail over five dynamic sessions. Thirty Thai students with low English proficiency underwent a “test–training–test” procedure using a specially designed CDRA program, which tested four reading skills: identifying the main idea, finding factual information, guessing word meaning, and making inferences. The results revealed a significant gain from pretest to posttest and steady improvement in actual scores from session to session.

Building on the findings obtained for receptive skills, the next logical step in the evolution of C-DA was its application to productive abilities. Of particular interest in this context is the work by Nicholas and Blake (2024) studying the pragmatics of English-language email writing through C-DA. Their two-tier system automatically identified specific pragmatic infelicities (e.g., inappropriate levels of request directness or missing closing formulas) and provided mediation at four levels of explicitness. The case study results of a Japanese student demonstrated uneven development: a rapid transition to self-regulation was observed in framing moves (openings and closings), whereas formulating the request required more explicit and prolonged support. Thus, C-DA made it possible to build a detailed diagnostic profile and simultaneously foster the development of pragmatic competence, confirming its effectiveness as a formative assessment tool in written discourse.

Computerized Dynamic Assessment and Grammar

Dynamic assessment has also been applied to the diagnosis of grammatical competence. One of the key studies specifically aimed at developing a C-DA tool for grammar assessment is the work of Modarresi and Alavi (2014), in which the Computerized Dynamic Grammar Test (CDGT) was created and validated. The grammar test covered 10 major structural categories (e.g. tenses, conditionals, modal verbs, prepositions, relative clauses) with a breakdown into 40 subcategories, thereby encompassing a wide range of grammatical phenomena. A fundamental methodological feature was the rejection of the multiple-choice format in favor of discourse-oriented tasks: students had to identify an error in a coherent text, which prevented guessing the answer as the graduated prompts (five levels – from most implicit to explicit) were presented. The results, obtained from a sample of 122 students, demonstrated a significant improvement in scores from non-mediated to mediated testing, and the Learning Potential Score (LPS) analysis confirmed that students with identical initial (non-mediated) scores differed substantially in their ability to use prompts.

Subsequently, Modarresi and Jeddy (2018) attempted to identify the relationship between computerized dynamic grammar assessment (CDGT) and fluid intelligence. The results showed that the correlation of the mediated score with fluid intelligence was strong, whereas the correlation of the non-mediated score with the same measure was only moderate, which, according to the authors, points to the more refined diagnostic sensitivity of DA.

The study by Behbahani and Karimpour (2025) addresses another specific grammatical form – reflexive pronouns – and raises the question of whether C-DA can promote the transition from explicit to implicit knowledge. In a quasi-experiment with two intact groups (N = 62, intermediate level), the experimental group underwent a C-DA procedure in which each mistake triggered graduated prompts (from implicit to explicit) embedded in interactive tasks; the control group was taught traditionally. Explicit knowledge was measured using the Untimed Grammaticality Judgment Test (UGJT), and implicit knowledge was measured through a sentence repetition task. The results showed a significant advantage for the experimental group in both explicit and implicit knowledge. Notably, no differences were found between learners with left- and right-hemisphere dominance, indicating the method's inclusivity.

Perhaps the most technologically advanced approach to C-DA of grammar is presented in the work of Jaganov, Blake, Villegas, and Carr (2025), who attempted to implement dynamic assessment of grammatical accuracy in free written production using large language models (LLMs). The DynaWrite platform, developed by the authors, provides four-level mediation – from an implicit indication of a mistake to an explicit explanation and the correct option – directly within the student's editing process of their own text. Comparative testing of 21 models revealed that GPT-4o was able to generate pedagogically appropriate feedback (adhering to sequence, gradation, and error resolution) in 68% of true positive detections. The system demonstrated acceptable response time and stability sufficient for real-time operation, and the financial costs proved negligible even for large-scale use. Thus, this work not only expands the scope of C-DA to productive writing skills but also opens a path to scalable, individualized dynamic assessment of grammar based on generative artificial intelligence.

A large-scale meta-analysis by Emadi and Arabmofrad (2025), covering 22 experimental and quasi-experimental studies from 2018–2024, confirms that, despite the acknowledged potential, grammar remains the least researched area in actual research practice: according to this meta-analysis, vocabulary and grammar attracted the least attention, accounting for only 4.5% of the selected studies.

Thus, the available empirical evidence confirms the effectiveness of C-DA for the development of grammatical competence; however, despite individual studies, systematic research specifically devoted to this aspect remains critically scarce.

Stage-Based Computerized Dynamic Assessment (SFS-LPS model)

The literature review indicates that C-DA can indeed measure learning potential (LPS) and has proven effective for developing grammatical competence. However, traditional C-DA typically does not capture the current stage of skill formation – a measure that is essential for understanding a learner’s developmental stage and for determining the most effective type of pedagogical support.

To address this need, the dynamic assessment methodology was complemented with a second, qualitative dimension: the Skill Formation Stage (SFS) (Akhrenova & Yatskina, 2026). This dimension is grounded in a language-pedagogy adaptation of Bernstein’s (1947) theory of the stage-based construction of skills, according to which any complex skill – including speech production – is formed in a sequential manner. Building on this framework, the concept of “stage of movement construction” was reinterpreted as a diagnosable Skill Formation Stage (SFS). In the proposed model, a skill is treated not as a binary category (acquired vs. not acquired) but as a hierarchy of stages, ranging from elementary reproduction (SFS-A) to flexible, creative use (SFS-E).

The operationalization of the model for grammatical skill is presented below (Akhrenova & Yatskina, 2026).

- *SFS-A (elementary reproduction)* – student recognizes the target grammatical structure among others, matches form and meaning according to a model, and performs actions by analogy.
- *SFS-B (algorithmic reproduction)* – student independently produces the grammatical form following an algorithm or template in an isolated context, relying on explicit markers.
- *SFS-C (contextual adaptation)* – student understands the difference in meaning conveyed by grammatical forms and selects the appropriate variant in a coherent text, guided by context and communicative purpose.
- *SFS-D (systemic integration)* – student uses various grammatical structures independently and appropriately in a cohesive utterance, combining them to achieve a communicative goal.
- *SFS-E (creative reproduction)* – student employs grammatical structures freely and creatively in spontaneous spoken and written production, choosing among forms to express subtle shades of meaning.

Thus, within this C-DA framework, the SFS-LPS model (Akhrenova & Yatskina, 2026) ensures all C-DA tasks in the digital platform are designed with reference to the SFS hierarchy – from exercises aimed at the basic formation of automaticity (SFS-A) to tasks requiring systemic integration of linguistic structures (SFS-D). The top level, SFS-E, which involves creative language use, is assessed by the teacher and is not included in the digital platform. The hierarchical design enables the system to diagnose not only the accuracy of response but also the qualitative stage of skill formation at which the learner is operating.

Research Questions

However, the empirical relationship between a learner’s dominant SFS and the specific type and amount of mediation they require - whether implicit prompts are sufficient or explicit guidance is necessary - has not been systematically examined. The present study addresses this gap by investigating the following research questions:

RQ1: How do learners with different dominant SFS (A-D) differ in the quantity (average number of hints) and the quality (distribution of implicit and explicit hints) of mediation required to complete tasks on their dominant level?

RQ2: What developmental patterns can be observed in the evolution of SFS and LPS across three C-DA cycles for learners with contrasting initial diagnostic profiles?

Methods

Design of the Study and Participants

The study adopts a multiple case study design (Yin, 2018), which is appropriate for exploring how learners with different SFS profiles respond to graduated mediation within a C-DA environment. Given the pilot nature of the investigation (N=16), the analysis is descriptive and context-sensitive, focusing on patterns observed within the present sample rather than statistical generalization.

All participants (N=16) were native speakers of Russian, had studied English as a compulsory school subject, and had attained a B2 level on the CEFR scale at the time of the study. None had an extended experience of living in an English-speaking country. Participation was voluntary and anonymous; informed consent was obtained from all participants.

Stage-Based C-DA Platform

The digital educational platform is designed to integrate individualized grammar instruction into a standard course curriculum through cyclic Stage-Based Dynamic Assessment. The proposed model follows a 70/30 framework, combining classroom instruction from a core textbook (70%) with personalized digital practice (30%). The platform's core algorithm conducts assessment across three cycles - before, during, and after a module - to measure students' learning potential (LPS) and grammar skill formation stage (SFS). Using these metrics, the system automatically generates personal assignments. As the system follows a cyclical design, ongoing monitoring and instructional adjustment are conducted over time.

SFS Diagnostic Tasks Design

The diagnostic tasks on the platform are organized according to the SFS hierarchy (A-D). When a learner makes a mistake, the system activates the mediation procedure, providing a series of hints: from the most implicit prompt (L1) to the most explicit (L4). If a learner does not respond correctly after L4, the item is recorded as unresolved (L5, Beyond ZPD). If a learner answers correctly on the first attempt, the item is recorded as L0. The progression of task types across SFS is summarized in Table 1.

As can be seen from Table 1, SFS-A tasks require recognition and identification of the target structure among distractors. Closed-ended formats (multiple choice, matching, sorting) are used, as the learner is not yet expected to produce the form independently. Mediation focuses on highlighting formal features and guiding identification. SFS-B tasks require the learner to apply the grammatical rule in isolated sentences following a clear algorithm. Formats include substitution, transformation, and gap-fill with explicit markers. Mediation activates the rule application sequence and reminds the learner of operational steps. SFS-C tasks require selecting between multiple grammatical forms in a coherent text (e.g., letters, posts, articles), where the choice is determined by communicative purpose. Mediation helps to analyse contextual cues and establishes links between form and meaning. SFS-D tasks are productive: learners create their own texts (micro-essays, e-mails, posts) on a given topic, using target structures

appropriately. Mediation prompts self-reflection and encourages expansion of the available repertoire.

Table 1.

SFS-framework Diagnostic Task Types

SFS	Task type	Example Format	Purpose
SFS-A (elementary reproduction)	Recognition, matching	Multiple choice, matching, sorting	Identify the grammatical form
SFS-B (algorithmic reproduction)	Substitution, transformation	Gap-fill, sentence transformation	Apply the rule algorithmically
SFS-C (contextual adaptation)	Context-based selection	Editing, cloze test	Distinguish meaning in context
SFS-D (systemic integration)	Productive writing	Essay, e-mail	Use forms in connected speech

Procedure

All participants completed the C-DA tasks in a computer-based environment. The platform administered tasks adaptively, starting from SFS-B and moving to higher or lower levels based on the learner's performance. For each learner, were recorded:

- the code of assistance (L0–L5) for each administered item;
- AS and MS for each item;
- LPS for each item (Calculated using the formula $LPS = (2*MS-AS)/MaxScore$ (Poehner, Zhang, & Lu, 2015);
- aggregated LPS for each SFS.

A detailed description of the procedure for all three C-DA cycles within a module and the principles for assigning individual tasks is provided in the paper (Akhrenova & Yatskina, 2026). For the purposes of this study, data from the first cycle were analyzed to address RQ1; data from all three cycles for two illustrative cases were analyzed to address RQ2.

Data Analysis

Based on the first cycle Stage-Based C-DA results, 16 participants were categorized according to their dominant SFS - the highest level at which the aggregated LPS reached the threshold of ≥ 1.0 . For each subgroup, the following metrics were calculated:

- mean AS and MS on the dominant level;
- average number of hints used per item on the dominant level;
- distribution of hint types: percentages of L0 (correct on first attempt), L1–L2 (implicit mediation), L3–L4 (explicit mediation), and L5 (Beyond ZPD) on the dominant level;
- mean aggregated LPS for each dominant SFS.

Given the case study design and the small sample size, the analysis is descriptive. No inferential statistical tests were applied. The findings are presented as patterns observed within the present

sample, with the intention of informing future large-scale validation research.

Results/Findings

Consistent with the multiple case study design of this pilot investigation, the analysis focuses on identifying patterns within the present sample (N=16) and exploring consistency with theoretical expectations, rather than on statistical generalization. Where relevant, multiple interpretations of the observed patterns are offered, and limitations are explicitly acknowledged.

Diagnostic Profiles: A Descriptive Overview

Table 2 presents the distribution of 16 participants across the four SFS levels, along with their mean Actual Scores (AS) and Mediated Scores (MS). All the metrics are calculated on items from the dominant SFS to ensure that the reported values reflect each subgroup's competence on the level that defines their diagnostic profile.

Table 2.

Distribution of Participants by Dominant SFS with Mean AS and MS

Dominant SFS	N	Mean AS	Mean MS	Δ (MS-AS)
SFS-A (elementary reproduction)	5	0.40	0.75	0.35
SFS-B (algorithmic reproduction)	7	0.21	0.59	0.38
SFS-C (contextual adaptation)	3	0.06	0.42	0.36
SFS-D (systemic integration)	1	0.50	0.90	0.40

Note. AS = Actual Score (proportion of items answered correctly on the first attempt); MS = Mediated Score (weighted by the level of assistance required, ranging from 0 to 1). All values are reported as proportions of the maximum score per item ($MaxScore=1$). The table includes only items from dominant SFS for each student. The SFS-D subgroup consists of a single participant, findings for this level are reported for completeness but are not generalizable.

As shown in Table 2, the mean MS is consistently higher than mean AS across all subgroups, confirming the effectiveness of the graduated mediation procedure within the present sample. The gain between MS and AS (Δ) is relatively stable across all SFS subgroups, ranging from 0.35 (SFS-A) to 0.40 (SFS-D). This stability tentatively suggests that, within this case study sample, learners at different stages of grammatical skill formation may require a comparable volume of support to bridge the gap between their independent and mediated performance. Given the pilot nature of the study (N=16) and the single participant in SFS-D, these findings should be interpreted as descriptive patterns requiring confirmation in a larger sample.

Mediation Patterns on the Dominant SFS

To examine the quantity and quality of mediation required, we analyzed the distribution of hint types used by learners on their dominant SFS during the first diagnostic cycle. Following the graduated mediation design, hints 1-2 were categorized as *implicit* (metacognitive and strategic prompts), while hints 3-4 were categorized as *explicit* (contextual and instructive prompts). Table 3 presents these metrics. The data are descriptive and are reported to illustrate how learners in different SFS subgroups responded to the graduated mediation procedure.

Table 3.

Mediation Efficiency Metrics on the Dominant SFS

Dominant SFS	N	Avg. Hints	% Implicit	% Explicit	% L0	% L5
SFS-A (elementary reproduction)	5	1.27	40.0%	13.3%	40.0%	6.7%
SFS-B (algorithmic reproduction)	7	2.04	46.4%	21.4%	21.4%	10.7%
SFS-C (contextual adaptation)	3	3.00	50.0%	22.2%	5.6%	27.8%
SFS-D (systemic integration)	1	0.50	50.0%	0.0%	50.0%	0.0%

Note. L0 = correct on the first attempt (no hints), L5 = Beyond ZPD (unresolved after all hints). All metrics are calculated on items from the dominant SFS for each learner. The SFS-D subgroup consists of a single participant; findings for this level are reported for completeness but are not generalizable.

Quantity of Mediation

The average number of hints used per item varies across subgroups: SFS-A (1.27), SFS-B (2.04), SFS-C (3.0), and SFS-D (0.50, $n=1$). The highest value is observed at SFS-C, indicating that learners at the Contextual Adaptation stage require the most intensive mediation. This pattern suggests that the quantity of support needed is not linearly related to SFS level but rather peaks at the intermediate stage, where learners are transitioning from algorithmic application to contextual use.

Quality of Mediation

Across the three subgroups with $n \geq 3$ (A, B, C), implicit hints account for 40.0% to 50.0% of all mediated responses. The predominance of implicit prompts across all levels is notable: even at the lowest level (SFS-A: 40.0%), learners respond to minimal, metacognitive guidance. The range of variation (10 percentage points) is relatively narrow, suggesting that the diagnostic algorithm consistently identifies levels at which learners are responsive to implicit support.

Explicit hints are used less frequently across all subgroups, ranging from 13.3% (SFS-A) to 22.2% (SFS-C). This pattern indicates that, on the dominant level, learners do not frequently require direct instruction. The absence of explicit prompts for the single SFS-D learner (0.0%) is consistent with this trend but cannot be generalized.

The proportion of items answered correctly without mediation is highest at SFS-A (40.0%) and SFS-D (50.0%, $n=1$), and lowest at SFS-C (5.6%). This U-shaped pattern reflects the nature of tasks at each level: reproductive tasks (A) are often solvable through rote recall, while productive tasks (D) reflect a high degree of internalization. Learners at the Contextual Adaptation stage (C), however, appear to be in a transitional zone where independent success is least frequent.

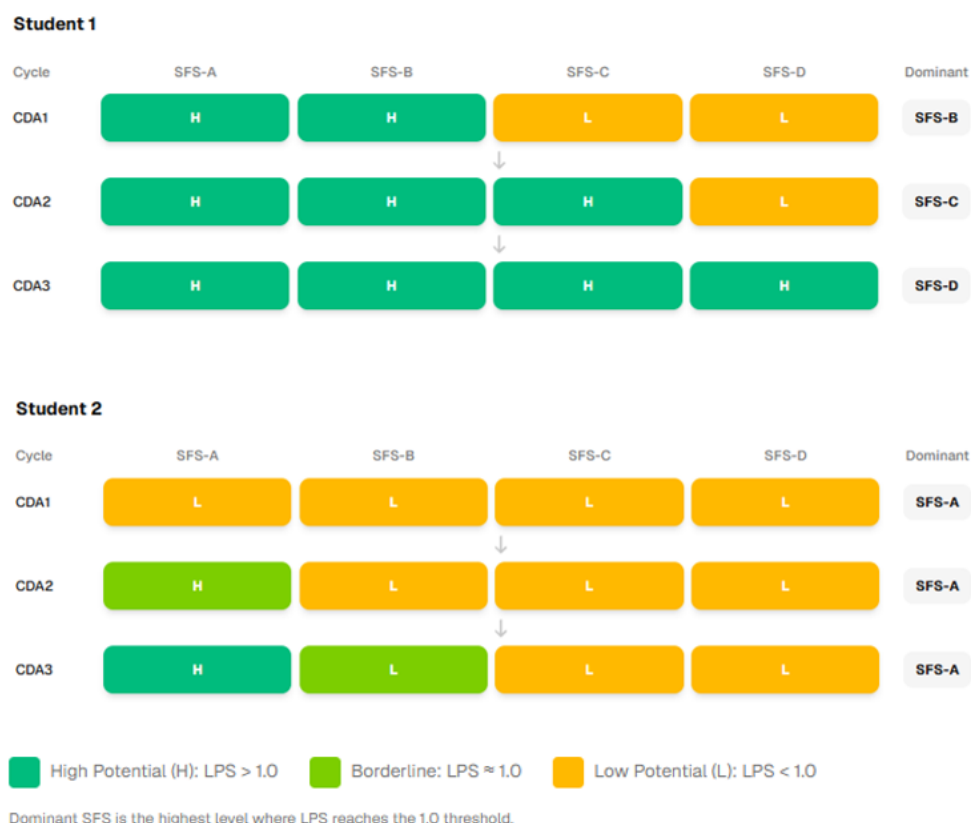
The rate of items that remained unresolved even after all hints were provided (Beyond ZPD, L5), is highest at SFS-C (27.8%), while other subgroups show substantially lower rates (SFS-A: 6.7%; SFS-B: 10.7%). This finding is particularly noteworthy, as it suggests that learners at the Contextual Adaptation stage frequently reach their cognitive limit even with full mediation. The concentration of L5 failures at SFS-C points to this level as a critical zone requiring further investigation.

Developmental Trajectories Across Cycles - Two Illustrative Cases

To illustrate the developmental potential of the SFS-LPS framework, two contrasting case studies selected from the sample are analyzed. Student 1 represents a case of notable progression from SFS-B to SFS-D across three cycles. Student 2 represents a case of stable plateau at SFS-A despite repeated mediation. Figure 1 visualizes the LPS dynamics and dominant SFS shifts for both learners across the three cycles.

Figure 1.

Heatmap of SFS-LPS Dynamics Across Cycles for Two Illustrative Cases



As shown in Figure 1, Student 1 demonstrates a clear upward trajectory across the three cycles. In Cycle 1, the dominant SFS is B (LPS = 1.2), with SFS-C showing borderline potential (LPS = 0.6). By Cycle 2, the dominant SFS shifts to C (LPS = 1.2), while SFS-D remains very low (LPS = 0.0). In Cycle 3, the dominant SFS reaches D (LPS = 1.4), with SFS-B and SFS-C maintaining high potential (LPS = 1.15 and 1.27, respectively). This progression through the SFS hierarchy, from B to C to D, indicates successful skill internalization across the three iterative cycles.

In contrast, Student 2 remains at SFS-A across all three cycles, illustrating a stable plateau. In Cycle 1, LPS at SFS-A is 0.4 (low). By Cycle 2, LPS rises to 0.9 (borderline), and in Cycle 3 it reaches 1.0 (high). Although the dominant SFS does not change formally, the increase in LPS from 0.4 to 1.0 indicates improved performance on the dominant level. This pattern suggests that, even without a formal shift in SFS classification, the system captures internal developmental changes through LPS dynamics.

These two contrasting cases illustrate the range of developmental trajectories captured by the SFS-LPS framework. Student 1's progression through the SFS hierarchy demonstrates the

system's capacity to support skill internalization over time. Student 2's plateau, despite increased LPS and reduced hint usage, highlights the system's diagnostic sensitivity in identifying learners who may require additional support.

Discussion

This case study set out to explore the diagnostic potential of the Stage-Based Computerized Dynamic Assessment (C-DA) platform for personalized grammar instruction by examining the relationship between learners' dominant Skill Formation Stage (SFS) and their mediation needs. The findings offer preliminary evidence that the SFS-based classification may differentiate learners according to both the quantity and quality of mediation required on their dominant level. Additionally, two illustrative case studies provide initial insights into the developmental trajectories captured by the framework across three iterative cycles, suggesting that the platform may offer a scalable, theoretically grounded tool for accommodating learner diversity in EFL classrooms.

The first research question (RQ1), which asked how learners with different dominant SFS levels differ in the quantity and quality of mediation required, was addressed through the analysis of average hint usage and hint type distribution. Within the present pilot sample ($N=16$), the average number of hints used per item showed a systematic pattern across SFS levels: SFS-A (1.27), SFS-B (2.04), SFS-C (3.00), and SFS-D (0.50, $n=1$). This pattern tentatively suggests that the quantity of mediation is not linearly related to SFS level but rather peaks at SFS-C, where learners appeared to require the most intensive support. The relatively low value at SFS-A may indicate that reproductive tasks are often solvable with minimal prompting, while the higher value at SFS-C suggests that the transition from algorithmic application to contextual use may represent a critical phase in skill development. This observation is supported by skill acquisition theories, which posit a non-linear trajectory in which learners may temporarily require more support as they navigate between developmental stages (Anderson, 1993).

Regarding the quality of mediation, the findings revealed that implicit hints (L1–L2) predominated across all SFS levels, accounting for 40.0% to 50.0% of mediated responses. This predominance is consistent with the theoretical expectation that learners operating within their Zone of Proximal Development (ZPD) should respond to minimal, metacognitive guidance (Aljaafreh & Lantolf, 1994). The fact that implicit prompts dominated even at SFS-A (40.0%) suggests that the diagnostic algorithm may successfully identify the level at which learners are most responsive to minimal support. Explicit hints (L3–L4) were used less frequently across all subgroups, ranging from 13.3% (SFS-A) to 22.2% (SFS-C). This pattern aligns with the logic of the SFS framework: on the dominant level, learners would not be expected to require frequent direct instruction, as this might indicate that the level has been overestimated.

The second research question (RQ2), which asked what developmental patterns can be observed in the evolution of SFS and LPS across cycles for learners with contrasting initial profiles, was addressed through two illustrative case studies. Student 1 demonstrated a clear upward trajectory, progressing from SFS-B in Cycle 1 to SFS-C in Cycle 2 and SFS-D in Cycle 3, with LPS values consistently above 1.0 on the dominant level. In contrast, Student 2 remained at SFS-A across all three cycles, although LPS increased from 0.4 to 1.0. This case highlights the diagnostic function of the SFS-LPS model: it can identify learners who may require additional supplementary support before they are ready to progress. The contrasting trajectories provide preliminary evidence that the iterative, cyclical design of the C-DA platform may support differentiated developmental pathways, with individual differences in responsiveness reflecting

differences in baseline ZPD.

Although these are promising findings, challenges and limitations have also been identified. The cross-sectional design for RQ1 compares different learners across SFS levels rather than tracking the same individuals over time; consequently, the observed differences should not be interpreted as evidence of developmental progression. The sample size (N=16) is modest, and subgroup sizes range from one to seven participants, limiting generalizability. The SFS-D subgroup consists of a single participant, and any conclusions regarding this level are tentative. The longitudinal case studies, while illustrative, are based on only two participants and should not be generalized. These limitations are consistent with the exploratory, pilot nature of this case study and should be addressed in future research.

The findings can also be applied to broader theoretical debates. One of them concerns the operationalization of the ZPD in digital learning environments. By showing that implicit prompts predominated across all SFS levels, and that explicit prompts increased only at the SFS-C, the findings illustrate how the SFS-LPS model might inform the design of graduated mediation that is sensitive to learners' current developmental stage. The two case studies further suggest that the framework may capture not only static diagnostic profiles but also developmental trajectories across cycles, demonstrating the potential of the iterative, cyclical design to support differentiated developmental pathways. This is particularly relevant for the broader goal of the platform: to offer a scalable, theoretically grounded tool for personalizing grammar instruction in diverse EFL classrooms, where learners at the same CEFR level may have vastly different mediation needs. This vision is consistent with research highlighting the growing importance of digital tools in supporting differentiated instruction, where teachers' roles are evolving towards facilitation and guidance rather than direct instruction (Ho, 2026).

Conclusion

In conclusion, this pilot case study has provided preliminary evidence that the SFS-based approach may improve the diagnosis of grammatical competence by differentiating learners according to their mediation needs, both in terms of quantity and quality of support. The two illustrative case studies further suggest that the framework may capture developmental trajectories over time, with one learner progressing through the SFS hierarchy and another remaining at the same level while demonstrating internal improvements in performance. The predominance of implicit mediation across all levels is consistent with theoretical expectations and provides initial empirical support for the diagnostic capacity of the framework. The effectiveness of the Stage-Based C-DA platform is connected to thoughtful integration, subsequent follow-up, and the recognition of its advantages and limitations as a diagnostic method. The platform thus offers a promising step towards personalized, scalable grammar instruction that accommodates learner diversity in mass education settings.

References

- Akhrenova, N. A., & Yatskina, K. V. (2026). Tsifrovaya model' diagnostiki obrazovatel'nykh traektorii v vysshem obrazovanii: Rezul'taty kvazieksperimental'nogo issledovaniya [Digital model for diagnosing learning trajectories in higher education: Results of a quasi-experimental study]. *Vysshee obrazovanie v Rossii* [Higher Education in Russia], 35(4), 126–145. <https://doi.org/10.31992/0869-3617-2026-35-4-126-145>
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language

- learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465–483. <https://doi.org/10.1111/j.1540-4781.1994.tb02064.x>
- Anderson, J. R. (1993). *Rules of the mind*. Psychology Press.
<https://doi.org/10.4324/9781315806938>
- Behbahani, H. K., & Karimpour, S. (2025). Exploring the impact of computerized dynamic assessment on the explicit and implicit knowledge of reflexive pronouns: The mediating role of brain dominance. *System*, 124, Article 103537.
<https://doi.org/10.1016/j.system.2024.103537>
- Bernshtein, N. A. (1947). *O postroenii dvizhenii* [On the construction of movements]. Medgiz.
- Burns, A. (2009). Grammar and communicative language teaching: Why, when and how to teach it? In *CamTESOL Conference on English Language Teaching: Selected Papers* (Vol. 5).
https://camtesol.org/Download/Earlier_Publications/Selected_Papers_Vol.5_2009.pdf
- Ebadi, S., Karimi, E., & Vakili, S. (2023). An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment. *Language Testing in Asia*, 13, Article 5. <https://doi.org/10.1186/s40468-023-00221-9>
- Emadi, M., & Arabmofrad, A. (2025). A meta-analytic study of computerized dynamic assessment in English language learning: Trends, effectiveness, and implications. *Journal of English Language Teaching and Learning*, 17(35), 117–136.
<https://doi.org/10.22034/elt.2025.65886.2756>
- Han, N. (2025). Language assessments in TESOL: Formative assessments and summative assessments. *SHS Web of Conferences*, 220, Article 04019.
<https://doi.org/10.1051/shsconf/202522004019>
- Ho, T. N. U. (2026). EFL teacher agency in the era of digital transformation: A multiple-case study in Vietnam. *International Journal of TESOL & Education*, 6(2), 19-29.
<https://doi.org/10.54855/ijte.26622>
- Jaganov, T., Blake, J., Villegas, J., & Carr, N. (2025). Large language model-driven dynamic assessment of grammatical accuracy in English language learner writing. *IEEE Access*. Advance online publication. <https://doi.org/10.1109/ACCESS.2025.3603191>
- Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. (2021). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 34(1-2), 1–30. <https://doi.org/10.1080/09588221.2019.1645181>
- Lantolf, J. P., & Poehner, M. E. (2010). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 11–33.
<https://doi.org/10.1177/1362168810383328>
- Meng, Y., Fu, H., & Wang, C. (2024). The effectiveness of computerized listening dynamic assessment: Attribute-based mediation model. *Language Learning & Technology*, 28(1), 1–28. <https://doi.org/10.64152/10125/73580>
- Modarresi, G., & Alavi, S. M. (2014). Designing and validating a test battery of computerized dynamic assessment of grammar. *TELL*, 8(2), 1–28.
<https://doi.org/10.22132/tel.2014.53816>
- Modarresi, G., & Jeddy, A. (2018). The association between dynamic assessment of grammar

- and fluid intelligence: A case of undergraduate EFL students. *International and Multidisciplinary Journal of Social Sciences*, 7(3), 297–321. <https://doi.org/10.17583/rimcis.2018.3881>
- Nasai, I. (2025). An analysis of the effectiveness of AI in learning English grammar in the digital era. *Jurnal Pendidikan Bumi Persada*, 4(1), 11–24. <https://jpbp.unbp.ac.id/index.php/jurnalpendidikan/article/view/76>
- Nicholas, A., & Blake, J. (2024). Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing. *Research Methods in Applied Linguistics*, 3(3), Article 100164. <https://doi.org/10.1016/j.rmal.2024.100164>
- Poehner, M. E., & Lantolf, J. P. (2021). The ZPD, second language learning, and the transposition ~ transformation dialectic. *Kul'turno-istoricheskaya psikhologiya* [Cultural-Historical Psychology], 17(3), 31–41. <https://doi.org/10.17759/chp.2021170306>
- Poehner, M., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337–357. <https://doi.org/10.1177/0265532214560390>
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge University Press.
- Qin, T., & Zhang, J. (2019). Computerized dynamic assessment and second language learning: Programmed mediation to promote future development. *Journal of Cognitive Education and Psychology*, 17(2), 198–213. <https://doi.org/10.1891/1945-8959.17.2.198>
- Schurz, A., & Coumel, M. (2023). Grammar teaching in ELT: A cross-national comparison of teacher-reported practices. *Language Teaching Research*, 27(5), 1167–1192. <https://doi.org/10.1177/1362168820964137>
- Siengyen, C., & Wasanasomsithi, P. (2024). Development of a computerized dynamic reading assessment program to measure English reading comprehension of Thai EFL undergraduate students. *rEFLections*, 31(3), 1215–1240. <https://so05.tci-thaijo.org/index.php/reflections/index>
- Thornbury, S. (1999). *How to teach grammar*. Pearson Education.
- Vo, T. Q. N., & Tran, K. M. A. (2025). The effectiveness of applying AI to online teaching English: from students and teachers' perspectives. *ICTE Conference Proceedings*, 9, 20–35. https://doi.org/10.54855/979-8-9870112-9-4_2
- Vygotsky, L. S. (1935). *Umstvennoe razvitie detei v protsesse obucheniya* [Intellectual development of children in the process of teaching/learning]. Uchpedgiz.
- Yarkofoji, S. I. (2019). Language skills and teaching of grammar. *International Journal of Strategic Research in Education, Technology and Humanities*, 6(2), 103–112. <https://internationalpolicybrief.org/wp-content/uploads/2023/10/ARTICLE10-83.pdf>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Biodata

Junior researcher, PhD candidate, RUDN University, Russia. Primary research interests focus on optimizing EFL group instruction and include methods of personalization and differentiation, as well as the practical application of core principles from educational psychology and pedagogy within contemporary technology-enhanced learning environments.