

The Adaptive Proficiency Acceleration Framework: A Diagnostic Driven Model for Individualizing Second Language Instruction

Raissa Enyonam Amele Amessouwoe¹, Kpoka Messanh Isaac², Julius Aderogba Odewabi³

¹ Wayne State University, Detroit, USA

amessouwoer@gmail.com

² Independent Researcher, USA

messanhkpoka24@gmail.com

³ University of Pittsburgh, USA

JAO166@pitt.edu

DOI: [10.56201/rjpst.vol.8.no8.2025.pg168.193](https://doi.org/10.56201/rjpst.vol.8.no8.2025.pg168.193)

Abstract

Knowing which component of a learner's second language competence is limiting progress is only half the problem; the other half is choosing, calibrating, and delivering an instructional response that actually fits the diagnosis. A teacher told that a learner has a working-memory-related fluency barrier gains little from that label alone unless it is translated into a specific technique, at a specific intensity, with a specific point at which support is withdrawn. This paper develops the Adaptive Proficiency Acceleration Framework, a model that converts a component-specific diagnostic signal into a matched instructional response, calibrates how much scaffolding that response should carry, and specifies when scaffolding should fade as proficiency accelerates. Grounded in Vygotskian mediation within the zone of proximal development, aptitude-treatment interaction research, comprehensible input and noticing theory, and the formative-assessment literature on actionable feedback, the framework organizes nine response types, from input modification and corrective feedback to technology-mediated adaptive practice and peer-mediated tasks, into a four-stage adaptive response cycle: signal interpretation, response selection, scaffolded delivery, and reassessment with fading. The paper states falsifiable propositions about which response types match which diagnostic signals, at what intensity, and specifies the evidence a future study would need to test whether matching a response to a diagnosis outperforms a generic, undifferentiated response. No intervention is reported here; the contribution is a decision architecture for individualizing instruction once a diagnosis exists, built so that a mismatch between response and diagnosis is itself detectable rather than absorbed into an unexplained null result.

Keywords

adaptive instructional response; zone of proximal development; aptitude-treatment interaction; scaffolding and fading; dynamic assessment; individualized second language instruction; comprehensible input; formative feedback; response calibration; second language pedagogy

1 Introduction

A diagnosis is only useful if someone can act on it. The applied linguistics literature has made considerable progress on the first half of individualized instruction, identifying which component of a learner's proficiency is limiting growth, through diagnostic testing, cognitive diagnostic modeling, and dynamic assessment (Isbell, 2021; 이연숙, 2014; Sun, Xu, & Wang, 2022). It has made less progress on the second half: once a specific barrier is identified, what instructional response actually fits it, how much of that response a given learner needs, and when the response should be withdrawn so it does not become a permanent crutch. A program that diagnoses meticulously and then responds with the same generic differentiation strategies regardless of diagnosis has not individualized instruction; it has only individualized the paperwork.

This paper develops the Adaptive Proficiency Acceleration Framework to close that gap. The framework assumes that a defensible diagnostic signal is available, whether from an instrument of the kind reviewed here or from a teacher's structured observation, and asks a different question from a purely diagnostic paper: given a specific, component-level signal, which instructional response is theoretically warranted, how intensely should it be delivered, and what evidence would show that the response is working well enough to begin withdrawing it. Literature coverage ends on 31 December 2025.

The case for treating response selection as its own problem, rather than as an obvious downstream consequence of diagnosis, rests on aptitude-treatment interaction research. DeKeyser (2019) argues that the same instructional treatment can help one learner and hinder another depending on that learner's aptitude profile, which means a response menu built around what is pedagogically fashionable, rather than around what the diagnostic and individual-differences literature actually supports for a given profile, will systematically help some learners while stalling others under the banner of individualization.

Four questions organize the paper. First, what response types does the second language instruction literature already support, and which diagnostic signal does each respond to? Second, how should the intensity of a response, the amount of scaffolding, repetition, or mediation it carries, be calibrated to a learner's profile rather than applied uniformly? Third, when and how should scaffolding fade as a targeted gain is achieved? Fourth, what evidence would show that a matched response outperforms an unmatched but otherwise similar one? These questions structure the theoretical review, the proposed model, the response taxonomy, and the research agenda that follow. Because no dataset is analyzed in this paper, every claim below is stated as a hypothesis, a decision rule, or a specification of required evidence, not as a finding. Where the paper illustrates a mismatch between response and diagnosis, the illustration is explicitly hypothetical rather than a report of an observed case, consistent with the paper's status as a conceptual and methodological contribution rather than an empirical one.

Three standards discipline the model throughout. Treatment validity requires that a response actually target the mechanism implicated by the diagnostic signal, not merely a construct that sounds related. Dosage validity requires that scaffolding intensity be specified and adjustable, rather than left to individual teacher intuition with no recorded rule. Fading validity requires an explicit criterion for withdrawing support, since a response that never fades is not accelerating a learner toward independent proficiency but substituting for it indefinitely.

The paper's contribution is the response architecture and its evidentiary requirements, not a claim of proven acceleration. Section 2 reviews the theoretical grounds for matching, calibrating, and fading instructional response. Section 3 proposes the four-stage adaptive response cycle. Section

4 specifies the review and evaluation methodology. Section 5 develops a response taxonomy across nine response types. Section 6 states falsifiable propositions about matching, intensity, and fading. Section 7 lays out a staged pathway for classroom adoption, and Sections 8 and 9 discuss implications and limitations.

The model is deliberately agnostic about which diagnostic instrument supplied the triggering signal, provided that instrument meets the construct-validity and specificity standards reviewed in Section 2.1. This is a design choice, not an oversight: a response architecture tightly coupled to one particular diagnostic instrument would inherit that instrument's limitations and would need to be rebuilt each time a program adopts a different or improved diagnostic tool, whereas an architecture built around the properties a signal must have, specificity, stability, and a stated observation window, remains usable across instruments that satisfy those properties.

2 Theoretical foundations for adaptive response

The theoretical anchor for scaffolded, individualized response is Vygotsky's (1978) account of the zone of proximal development, the gap between what a learner can do unassisted and what the same learner can do with mediation. Dynamic assessment operationalizes this gap directly, using the size and type of mediation a learner needs as the diagnostic and instructional signal at once (Sun, Xu, & Wang, 2022), and Ghanbarpour's (2017) meta-synthesis finds that mediated-performance gaps predict subsequent independent gains more reliably than static scores, which is the theoretical basis for treating scaffolding intensity, not just presence or absence of instruction, as something to calibrate rather than standardize.

Comprehensible input theory supplies a second mechanism relevant to response selection. Krashen (1985) argues that acquisition depends on input calibrated just beyond a learner's current level, a claim with a direct implication for adaptive response: input-modification techniques, simplification, elaboration, and paced delivery, should be calibrated to the diagnosed level of the specific skill they target rather than to a learner's global proficiency band. Schmidt (1990) adds that input becomes intake only when a learner notices the relevant feature, which is why several response types developed in Section 5 pair input modification with an explicit attention-directing device rather than relying on exposure alone. This calibration principle is consistent with de Bot's (2014) finding that early but low-quality exposure does not reliably outperform later, higher-quality exposure, and Navracics's (2014) finding that the communicative intimacy of input predicts outcomes more than raw quantity, both of which support calibrating input-modification intensity to interactional quality rather than volume alone. Formative feedback research supplies the theoretical basis for corrective and noticing-oriented response types specifically. Hattie and Timperley (2007) show that feedback effectiveness depends on specificity, timing, and whether it answers where a learner is going, how they are doing, and what to do next, and Black and Wiliam (1998) establish the broader principle that feedback functions formatively only when it changes subsequent teaching or learning action. Ketabi and Ketabi (2014) apply this logic specifically to second and foreign language classrooms, and Lee and Coniam (2013) show that assessment-for-learning practices in EFL writing classrooms depend on feedback specific enough for a learner to act on, which is the same specificity requirement the response taxonomy in Section 5 applies to every response type, not only feedback-based ones.

Aptitude-treatment interaction research, reviewed by DeKeyser (2019), establishes that response selection cannot be separated from the learner's cognitive and affective profile without risk. Shin

and Christianson's (2011) finding that structural priming in second language learners depends on working memory capacity, and Wakabayashi's (2009) and Pulido's (2003) work on how processing capacity and topic familiarity shape lexical learning, together imply that a response calibrated for a high-working-memory learner may be poorly calibrated, not merely less optimal, for a learner with a genuine working-memory barrier, a distinction the intensity-calibration stage of the model in Section 3 is built to capture. This is the central reason response selection is modeled here as a distinct stage rather than a fixed lookup from signal to technique: the same signal can warrant different responses depending on aptitude, and the same response can warrant different intensities depending on the same variable.

Motivational and affective theory supplies the grounds for treating some responses as regulating engagement rather than directly building linguistic knowledge. Gardner's (1985) socio-educational model and Dornyei's (2005) synthesis of the psychology of the language learner both treat motivation quality and anxiety as responsive to instructional design, not fixed, and Nida (1956) and Spolsky (1969) establish the earlier case that affective variables can be the operative barrier a response should target directly. Foster-Cohen (2004) cautions against assuming affective response always operates through linguistic mechanisms, supporting the taxonomy's decision to treat affective-regulation responses as a distinct response type in Section 5 rather than folding them into linguistic-skill responses.

Technology-mediated delivery is the final theoretical strand. Adaptive e-learning models (Kostolanyova & Simonova, 2020a, 2020b) and adaptive self-assessment within massive open online course environments (Hashim, Salam, Nurul, & Syafiatun, 2018) show that response sequencing and intensity can, in principle, be automated, and Menaka and Sankar (2019) report broader technology-mediated assessment and response practice for second language learners. Li's (2024) more recent examination of generative-model writing feedback shows both the promise and the limits of this automation, finding that feedback quality varies by error type in ways a response protocol must anticipate; automation changes how a response is delivered, not whether it still needs to be matched to a diagnosis. Dunkel (1999) anticipated this promise early in computer-adaptive testing research, and Forgor (2021) found that unequal digital access alone explained a meaningful share of an observed proficiency gap in one assessment-for-learning study, a reminder that technology-mediated response depends on infrastructure equity as much as on algorithm design. This technology-mediated strand is not confined to language learning. Onwubuya, Fawehinmi, Sunday, Igwenagu, Adeniran, and Agyapong (2026) report parallel gains from artificial-intelligence-powered adaptive learning systems in STEM education, evidence that automated response sequencing is maturing as a general instructional-technology capability whose diagnostic-to-response logic the present framework adapts to the specific error taxonomy of second-language proficiency rather than importing wholesale.

Reporting outcomes at the level of a proficiency scale, rather than only at the level of an individual construct, gives the model a common vocabulary for communicating progress to learners, parents, and receiving institutions. Xu and Li (2024) provide practical guidance for CEFR-informed reporting that this paper adopts for that purpose specifically: a response type's success is reported both as movement on the targeted construct and, where a CEFR-referenced descriptor exists for that construct, as movement relative to that descriptor, so a gain recorded internally by the model remains interpretable to an audience outside the program that produced it. Reporting at scale level also supports the classroom-wide adoption decision in Section 7: a matching advantage large enough to move a learner's classification on a recognized proficiency

scale is easier to weigh against training and delivery cost than a matching advantage visible only as a change in an internal, program-specific score.

Taken together, this literature establishes mechanisms for matching a response to a diagnosis, calibrating its intensity, and fading it, but it does not integrate these mechanisms into a single decision architecture that a teacher or program can actually apply. Oxford's (2017) treatment of learning strategies and Ellis's (2008) broad synthesis of second language acquisition both gesture toward this kind of integration without formalizing the matching, calibration, and fading decisions explicitly. The model proposed in the next section formalizes them.

2.1 The diagnostic starting point: signal quality and its limits

Before a response can be selected, the signal that triggers selection must itself be trustworthy, and not every instrument that reports a language score is built to support that trust. Alderson's (2005) distinction between a diagnostic test and a proficiency test, and Bachman and Palmer's (2010) construct-first approach to assessment, both warn that an instrument optimized for ranking learners against each other, the kind the Common European Framework of Reference is often used for in practice even though it was designed to support finer-grained profiling (Council of Europe, 2020), can produce a score that looks componential without actually licensing a specific response. Fulcher (2010) makes the same point from a practical testing perspective: a test relabeled as diagnostic without a validation argument for a specific pedagogical inference is not thereby diagnostic. Harding, Alderson, and Brunfaut's (2015) diagnostic principles for reading and listening, and Thomas's (1994) review of proficiency assessment in acquisition research, both supply criteria this paper adopts for judging whether an upstream signal is strong enough to drive response selection rather than only description.

Signal specificity varies by skill. Yang's (2021) diagnostic reading instrument for beginning Chinese-as-a-second-language learners separates decoding from comprehension difficulty explicitly, which means a signal from that kind of instrument can safely drive the choice between an input-modification response and a strategy-instruction response; a global reading score cannot make that distinction and should not be treated as though it can. Figueiredo, Martins, and Fernandes da Silva (2013) show that writing instruments calibrated on one literacy population misclassify learners from a different background, a caution the response-selection stage in Section 3 addresses by requiring the aptitude-treatment cross-check before a response is finalized. Johnson (1992) and Teló, Kivistö de Souza, O'Brien, and Carlet (2024) both show that the modality in which a productive skill is assessed, written versus auditory, self-assessed versus rater-assessed, changes the picture that results, so a response model that treats a diagnostic signal as modality-neutral risks selecting a response for the wrong modality. Waniek-Klimczak's (2018) review of pronunciation assessment research reinforces this point specifically for the intelligibility-accentedness distinction central to the corrective-feedback response type in Section 5. Lexicogrammatical signals require particular care because breadth, depth, and grammatical control are frequently reported as a single vocabulary or grammar score despite measuring different things (Nation, 2001; In'nami, 2013; Koizumi, 2015). Spinner (2011) shows that morphosyntactic subskills develop unevenly within a learner, which means a grammar-focused response selected on an aggregate grammar score risks targeting a subskill the learner has already mastered while leaving the actual barrier unaddressed. A first-language transfer interpretation of a given error pattern must additionally be checked against a developmental-universals alternative before it drives response selection (Jakobovits, 1969; Gass, 1989; Sang,

1983), since a contrastive response selected for a transfer-attributed error that is actually developmental will target the wrong mechanism, a risk this paper inherits directly from the diagnostic-stage literature it builds on.

A signal that is biased before it reaches the response-selection stage will bias the response that follows, so the response model inherits the fairness requirements the diagnostic literature has already established. Roever (2007) shows that differential item functioning in pragmatics assessment can make a pragmatic barrier look like a general proficiency effect, which would misdirect response selection toward a linguistic-skill response rather than the pragmatic-competence-specific response the taxonomy in Section 5 actually provides. Badrun Kholid, Rahman, and Irawan (2024) and Yeotak Yoon, 김정우, and Min (2008) both report that diagnostic instruments calibrated in one classroom context produced different classification patterns in another, reinforcing that the response-selection stage cannot simply trust an imported signal without a local check. Self-report signals, useful for constructs such as willingness to communicate where the learner is the most direct source of evidence, are less reliable for externally observable constructs (Butler, 2023, 2024), a distinction that determines which response types in Section 5 can safely be triggered by self-report alone versus which require an external or performance-based signal. Zhang (2013) and Lantaigne (2017) both address writing-specific signal design, Zhang arguing that writing functions simultaneously as a target skill and a vehicle for learning, and Lantaigne proposing sentence-reordering tasks that isolate writing-specific mechanics from broader grammatical knowledge, a separation the written-production-adjacent response types in Section 5 depend on to avoid targeting the wrong mechanism. Instrument choice also interacts with age and setting in ways the response model must anticipate rather than assume away. A signal drawn from a self-report instrument may be usable for an older, metacognitively mature postsecondary learner in a way it is not for a younger secondary-level learner who has not yet developed the self-monitoring capacity self-report presupposes, and a signal drawn from a computer-adaptive instrument presupposes a level of digital access and familiarity that cannot be assumed uniformly across settings (Forgor, 2021). The response-selection stage in Section 3 therefore treats the appropriateness of the signal-generating instrument itself, not only the signal's content, as part of what must be checked before a response is finalized. None of this instrumentation literature was developed to answer the question this paper asks, but each strand constrains what the response-selection stage in Section 3 is entitled to assume about the signal it receives. A response model built without attention to upstream signal quality would simply relocate the diagnostic field's unsolved problems into the response layer instead of resolving them, which is why Section 3's signal-interpretation stage is written as an active classification step rather than a passive intake of whatever score arrives.

3 The adaptive proficiency acceleration model

The Adaptive Proficiency Acceleration Framework organizes individualized response across four linked stages: signal interpretation, response selection, scaffolded delivery, and reassessment with fading. Unlike a one-time prescription, the cycle is designed to repeat, with each pass-through reassessment feeding a revised signal back into the next round of response selection, so that a response calibrated for a learner's profile today is not assumed to remain optimal indefinitely. Signal interpretation takes a component-specific diagnostic signal, of the kind a diagnostic instrument or structured teacher observation can provide, and classifies it along two dimensions before any response is selected: whether the barrier appears linguistic, cognitive,

or affective in character, and whether it appears stable or transient. This classification determines which region of the response taxonomy in Section 5 is even eligible, since, for example, an affective barrier misclassified as linguistic will draw a response, such as further grammar drilling, that targets the wrong mechanism entirely.

Response selection matches the classified signal to one or more response types from the taxonomy, using the aptitude-treatment interaction logic reviewed in Section 2 to check the match against what is known about the learner's broader cognitive and affective profile, not only the presenting signal. A response selected on signal alone, without this cross-check, risks the mismatch DeKeyser (2019) documents, in which a treatment that helps most learners with a given profile actively hinders a learner whose broader profile differs from that norm.

Scaffolded delivery calibrates intensity, the amount of mediation, repetition, or simplification the response carries, using the zone-of-proximal-development logic from Vygotsky (1978) and dynamic assessment's operationalization of it (Sun, Xu, & Wang, 2022): intensity should track the size of the gap between unassisted and mediated performance on the targeted construct, not a fixed program-wide default. Delivery fidelity, whether the response was actually delivered at the calibrated intensity, is recorded here so that a disappointing outcome can be traced to a mismatch, an intensity error, or a genuine non-response rather than defaulting to the least informative explanation. Reassessment with fading closes the cycle. The targeted construct is reassessed on the same terms as the original diagnostic signal, scaffolding intensity is reduced according to a predetermined fading rule when the gap between assisted and unassisted performance narrows, and the revised signal, including any newly visible barrier that the original response has uncovered, becomes the input to the next pass through response selection. A response that shows no narrowing after a defined number of cycles should trigger re-classification at the signal-interpretation stage rather than simple repetition of the same response at higher intensity.

Table 1. The adaptive response cycle: four stages

Stage	Function	Key mechanism	Decision it supports
Signal interpretation	Classify a diagnostic signal before any response is chosen	Linguistic, cognitive, or affective character; stability over time	Determine which region of the response taxonomy is eligible
Response selection	Match a response type to the classified signal	Aptitude-treatment interaction cross-check against the learner's broader profile	Avoid a response that fits the signal but not the learner
Scaffolded delivery	Calibrate and deliver scaffolding intensity	Zone-of-proximal-development gap between assisted and unassisted performance	Set and record the intensity actually delivered
Reassessment and fading	Re-measure the target and adjust intensity	Predetermined fading rule tied to narrowing of the assisted-unassisted gap	Continue, taper, or re-classify the signal

3.1 A hypothetical illustration of the cycle

The following illustration is offered to clarify how the four stages interact; it is a hypothetical walkthrough, not a reported case, and no learner data underlie it. Consider a learner diagnosed, through an instrument of the kind reviewed in Section 2.1, with a component-specific barrier in spoken fluency co-occurring with elevated language anxiety and otherwise unremarkable grammatical and lexical knowledge. At signal interpretation, this profile is classified as primarily affective with a secondary cognitive component, since the fluency gap appears concentrated in higher-stakes interactional contexts rather than in low-stakes production, a pattern the classification rules in Section 3 treat as evidence against a purely linguistic barrier.

At response selection, this classification rules out corrective feedback and input modification as first-line responses, since neither targets an affective mechanism, and instead selects affective regulation as the primary response with scaffolded production as a secondary response once anxiety begins to decline. The aptitude-treatment cross-check confirms this selection is consistent with the learner's broader profile: nothing in the working-memory or processing-speed domain suggests a cognitive-capacity barrier that scaffolded production alone, without affective regulation first, would need to address.

At scaffolded delivery, affective regulation is delivered at an intensity calibrated to the size of the gap between the learner's low-stakes and high-stakes performance, starting with structured, low-audience practice and a stated criterion for increasing audience size, consistent with the intensity rule specified for this response type in Section 5.6. Delivery fidelity is recorded by logging actual audience size and frequency against the calibrated plan, so a disappointing outcome can later be distinguished from a delivery shortfall.

At reassessment with fading, the same fluency and anxiety measures used at signal interpretation are re-administered. If the gap between low-stakes and high-stakes performance has narrowed, audience size and stakes are increased according to the fading rule, and scaffolded production is introduced as the secondary response now that the affective barrier is receding. If the gap has not narrowed after the defined number of cycles specified in Section 3, the model calls for re-classification at signal interpretation rather than simply repeating affective regulation at higher intensity, since a persistent gap after appropriately dosed affective regulation is itself evidence that the original classification may have been incomplete, for example, that a genuine cognitive-processing barrier was present alongside the affective one from the start.

This illustration is deliberately drawn to show a case where response selection is not obvious from the presenting signal alone, precisely because the easy cases, a clear grammatical error pattern met with corrective feedback, do not illustrate why the four-stage cycle, rather than a direct signal-to-response lookup table, is necessary. The harder cases, where an affective or contextual condition shapes how a linguistic signal should be read, are where the model's distinction between signal interpretation and response selection earns its place in the architecture.

3.2 Evidence synthesis by response mechanism

Five mechanisms recur across the individualized-instruction literature and organize how the response taxonomy in Section 5 was built, each supplying a different lever a response can pull.

Mediation-based scaffolding, the mechanism behind dynamic assessment, works by supplying exactly the assistance a learner needs and no more, then withdrawing it as competence grows (Sun, Xu, & Wang, 2022; Ghanbarpour, 2017). Mehri and Amerian (2015) catalogue the practical cost of this mechanism, including the time required for one-on-one mediation and the

training mediators need to calibrate assistance consistently, a cost the intensity-calibration stage of the model must weigh against the mechanism's precision.

Input and attention mechanisms work by adjusting what a learner is exposed to and directing attention to a specific feature within it (Krashen, 1985; Schmidt, 1990). Gaillard and Tremblay's (2016) elicited imitation research shows that learners' real-time processing of modified input can be measured directly, which gives this response family one of the few mechanisms in the taxonomy with a built-in, process-level check on whether the modified input is actually being processed differently, rather than only whether an outcome score eventually moves.

Feedback and noticing mechanisms work by making an error or gap visible to the learner at a point where the learner can act on it (Hattie & Timperley, 2007; Black & Wiliam, 1998). Neumann (2014) shows that teacher-generated feedback on grammatical ability is prone to systematic bias, which is why the response taxonomy pairs feedback-based responses with a structured protocol, such as the checklist Safari and Ahmadi (2023) validate for writing, rather than relying on unstructured teacher impression, and Oh (2019) shows that the resources a learner draws on to act on feedback, including translation and first-language planning, shape whether the feedback mechanism actually changes subsequent performance.

Strategy and metacognitive mechanisms work by building the learner's own capacity to monitor, plan, and compensate, rather than by directly supplying linguistic knowledge (Oxford, 2017). Koda (2007) shows that crosslinguistic differences in metacognitive monitoring partly explain differences in reading development, which argues for treating strategy instruction as a response type in its own right rather than an add-on to other responses, and Macaro (2003) and Brown (2013) both argue that classroom teachers are well placed to deliver this kind of response themselves once it is specified clearly enough to teach. Caplan's (2019) argument that language is learned through language, not only about language, supports embedding strategy instruction within meaningful communicative tasks rather than teaching strategies as an isolated skill disconnected from content.

Affective-regulation mechanisms work by reducing anxiety or building self-efficacy and willingness to communicate so that a learner's existing competence becomes visible under classroom conditions (Gardner, 1985; Nida, 1956; Spolsky, 1969). Heriansyah, Darni, Fajri, and Sahardin (2021) show that curriculum and assessment design itself shapes how much affective barrier a given classroom produces, which is why the response taxonomy treats affective-regulation responses as targeting a barrier that is partly instructional in origin, not purely a fixed trait of the learner.

4 Research design and evaluation

The review searched Linguistics and Language Behavior Abstracts, ERIC, Scopus, and Google Scholar for English- and French-language peer-reviewed work on individualized and differentiated second language instruction, dynamic and formative assessment, aptitude-treatment interaction, and technology-mediated adaptive delivery, published through 31 December 2025. Search terms combined each of the nine response types developed in Section 5 with terms for scaffolding, feedback, and adaptation; citation chaining from key theoretical sources (Vygotsky, 1978; Krashen, 1985; DeKeyser, 2019) identified additional foundational work. Appraisal distinguished studies that report evidence for a specific response mechanism, of the kind that can inform the matching and intensity rules in Section 5, from studies that report an intervention's overall effect without isolating which mechanism drove it. Only the former were

treated as direct evidence for the response taxonomy's matching logic, while broader intervention studies informed the theoretical synthesis in Section 2 without being treated as evidence for a specific match. Because response validation depends on comparing matched against unmatched assignment, statistical power considerations differ from those of a purely diagnostic validation study: the sample must be large enough to detect an interaction, not only a main effect, which typically requires a larger sample than a simple response-versus-no-response comparison of the same nominal size. Future studies should power their design for the interaction specified in the relevant proposition in Section 6, not for a main effect that the proposition does not actually claim. Programs planning such a study should budget accordingly: an underpowered interaction test is likely to return a null result regardless of whether a genuine matching advantage exists, and that null result would be reported, under the standards in Section 4, as inconclusive rather than as evidence against the taxonomy. Response types in Section 5 were developed by extracting, from each mechanism-specific study identified in the review, the diagnostic signal the mechanism is theoretically suited to address, then grouping techniques that share a mechanism into a response type. Each response type retains a stated target signal, an intensity-calibration rule, and a fading criterion, so the taxonomy can be audited against its theoretical sources rather than treated as a menu of generically desirable practices. Future validation requires studies that randomize or otherwise counterbalance response assignment across learners with a defined diagnostic signal, compare a matched response against an unmatched but otherwise similar one, report intensity actually delivered rather than intensity intended, and track fading, not only immediate post-response gains. A response type that does not outperform a well-delivered unmatched alternative in this design should be revised or removed from the taxonomy rather than retained on theoretical grounds alone. Because this model concerns instructional response rather than diagnostic classification, ethical review of any future empirical test must additionally examine the risk of withholding a response from a comparison group whose diagnosed barrier is genuine; a matched-versus-unmatched design, rather than a response-versus-no-response design, is the ethically preferable way to test the propositions in Section 6 for exactly this reason. Where a comparison group's diagnosed barrier is severe enough that any delay in a targeted response carries meaningful risk, the comparison response should still be an active, theoretically motivated alternative rather than an inert placebo, consistent with standard practice in educational and clinical intervention research more broadly.

A further methodological limitation concerns the review's scope. Restricting the search to English- and French-language, peer-reviewed sources through 31 December 2025 likely underrepresents response-mechanism research published in other languages or in practitioner-oriented outlets outside the databases searched, and the nine response types developed in Section 5 should be read as a synthesis of what this literature currently supports rather than an exhaustive account of every response mechanism a future review might surface.

Table 2. Evaluation criteria for the adaptive response model

Criterion	What it requires of a response type	Interpretation if the criterion fails
Treatment validity	Evidence that the response acts on the mechanism implicated by the target signal, not a related but distinct one	The response may still help for reasons unrelated to the stated diagnosis

Dosage validity	A stated, adjustable intensity rule tied to the assisted-unassisted performance gap	Intensity was left to unrecorded teacher judgment rather than calibrated
Fading validity	An explicit criterion for withdrawing scaffolding as the gap narrows	The response risks becoming a permanent substitute rather than an accelerant
Delivery fidelity	Evidence that the intended intensity was actually delivered	A disappointing outcome may reflect delivery failure rather than a bad match
Matching advantage	Comparison of a matched response against an unmatched but otherwise similar one	Any observed gain may reflect general instructional attention rather than matching
Transfer of gain	Evidence that a gain achieved under scaffolding persists after fading	Gains may be scaffold-dependent rather than genuinely accelerated
Equity of response	Evidence that response assignment does not systematically disadvantage a learner subgroup	A matching rule may be encoding bias rather than genuine treatment interaction

5 Response taxonomy and intensity matrix

The taxonomy organizes nine response types, each keyed to the mechanism it draws on, the diagnostic signal it is theoretically suited to address, and a stated intensity-calibration and fading rule. The response types are not mutually exclusive; a single diagnosed barrier can draw on more than one, and the response-selection stage in Section 3 is where a specific combination is chosen for a specific learner. Each subsection below follows the same pattern: what mechanism the response type draws on, which diagnostic signal it targets, what intensity means concretely for that type, and what fading looks like once the targeted gap narrows, so the nine descriptions can be compared directly against one another and against the summary matrix in Table 3.

5.1 Input-modification responses

Input-modification responses adjust the linguistic complexity, pacing, or elaboration of the material a learner receives, targeting signals in receptive processing such as listening or reading comprehension gaps that persist despite intact decoding. Krashen's (1985) comprehensible-input logic sets the calibration target just beyond the learner's current level rather than at a fixed grade-level default, and Mislevy and Yin's (2009) complex-adaptive-systems view of language cautions that input modified on one dimension, such as vocabulary frequency, can inadvertently change another, such as syntactic complexity, so a genuinely calibrated input-modification response specifies which dimension is being adjusted and monitors the others for unintended drift. Intensity here means the size of the gap between input level and learner level; fading means narrowing that gap toward unmodified input as comprehension stabilizes.

5.2 Output-and-interaction responses

Output-and-interaction responses require learners to produce language under communicative pressure, targeting signals in spoken fluency, interactional competence, and grammatical control that show up more reliably in production than in comprehension. Gaillard and Tremblay's (2016)

elicited imitation research demonstrates that structured production tasks can isolate real-time processing difficulty from knowledge gaps, a distinction that matters for calibrating this response type, since a learner who knows a structure but cannot deploy it under time pressure needs a different intensity profile than a learner who has not yet acquired the structure at all. Mohamadi (2017) shows that task engagement itself is a legitimate object of assessment and, by extension, of response design, not merely a precondition for it, which supports treating task structure as an active ingredient of this response type rather than a neutral delivery vehicle for content. Intensity here means the complexity and time pressure of the production task; fading means increasing task complexity and reducing planning time as fluency stabilizes.

5.3 Corrective feedback and noticing responses

Corrective feedback and noticing responses make a specific error or gap visible at a point where the learner can act on it, targeting signals in grammatical control, pronunciation, and writing accuracy where the underlying knowledge may already be partly present but not yet consistently applied. Schmidt's (1990) noticing hypothesis specifies the mechanism directly: feedback works by directing attention to a mismatch between the learner's output and the target form, and Hattie and Timperley (2007) add that feedback effectiveness depends on answering where the learner is going, how they are doing, and what to do next, not on volume of correction. Intensity here means feedback specificity and frequency; fading means moving from explicit correction toward learner self-monitoring as the targeted error becomes less frequent. Because the same surface error can arise from a transfer pattern or a developmental sequence, as Section 2.1 discusses, feedback content itself should differ, contrastive explanation in the former case and repeated exposure to the target form in context in the latter, even when the response type and intensity rule are otherwise identical.

5.4 Scaffolded production responses

Scaffolded production responses supply graduated mediation within a learner's zone of proximal development, targeting signals where a learner's assisted performance substantially exceeds unassisted performance, which dynamic assessment is built to detect (Vygotsky, 1978; Sun, Xu, & Wang, 2022). Ghanbarpour's (2017) synthesis shows this response type is among the most consistently effective in the reviewed literature, but Mehri and Amerian (2015) show it is also among the most resource-intensive to deliver with fidelity, so the taxonomy pairs it with an explicit resourcing note in the implementation pathway in Section 7. Intensity here means the type and amount of mediation, from a prompt to a full model; fading means moving up a graduated mediation hierarchy toward independent performance as the assisted-unassisted gap narrows.

5.5 Strategy-instruction responses

Strategy-instruction responses build a learner's own capacity to monitor, plan, and compensate for a linguistic gap, targeting signals in strategic competence and metacognitive control rather than a specific linguistic subsystem. Oxford (2017) catalogues the strategy types this response draws on, and Koda's (2007) finding that metacognitive monitoring differences partly explain crosslinguistic reading-development differences supports treating this response as building a transferable capacity rather than a single-skill fix. Intensity here means how much of the strategy is modeled versus practiced independently; fading means shifting from teacher-modeled to

learner-initiated strategy use.

5.6 Affective-regulation responses

Affective-regulation responses target language anxiety, low self-efficacy, or low willingness to communicate directly, on the theory that these barriers suppress the expression of competence that other response types are working to build. Gardner's (1985) distinction between integrative and instrumental motivation and Foster-Cohen's (2004) caution against assuming affective variables reduce to linguistic mechanisms both support treating this response type as targeting a genuinely separate mechanism rather than a byproduct of linguistic progress. Intensity here means the structure and frequency of low-stakes practice opportunities; fading means increasing stakes and audience size as measured anxiety declines and willingness to communicate rises.

5.7 Technology-mediated adaptive responses

Technology-mediated adaptive responses use computer-adaptive sequencing to adjust task difficulty and practice frequency automatically, targeting signals across multiple domains where practice volume, not mediation quality, is the binding constraint. Kostolanyova and Simonova's (2020a, 2020b) adaptive e-learning models and Hashim, Salam, Nurul, and Syafiatun's (2018) adaptive self-assessment platforms illustrate the mechanism, but Norris's (2001) concerns about automated oral-proficiency scoring and Li's (2024) finding that generative feedback quality varies by error type both argue for human review of automated response decisions rather than full delegation. Intensity here means practice frequency and algorithmic difficulty step size; fading means the system reducing practice frequency as accuracy and speed stabilize.

5.8 Peer-mediated and collaborative responses

Peer-mediated and collaborative responses use structured interaction with another learner, rather than only teacher mediation, to supply scaffolding, targeting signals in interactional competence and willingness to communicate where teacher-learner interaction alone provides limited practice volume. Rafi and Pourdana (2023) find that collaborative oral tasks within a differentiated framework surface barriers that neither individual performance sampling nor teacher-only mediation reveals, and Poole's (1992) research on classroom language socialization cautions that peer interaction structures can themselves position some learners as more or less competent, which the response taxonomy addresses by requiring structured, rotating pairing rather than fixed groupings. Intensity here means the structure and frequency of peer tasks; fading means increasing task openness as interactional competence stabilizes.

5.9 Fading and transfer responses

Fading and transfer responses are not a separate technique so much as the closing mechanism every other response type in this taxonomy must specify. Pourdana's (2022) finding that short-term gains from targeted intervention did not always predict durable improvement is the direct empirical warrant for treating fading and transfer as their own response category rather than an afterthought, and Ströbel, Kerz, and Wiechmann's (2020) finding that first-language writing complexity predicts second-language writing complexity in advanced learners suggests that transfer checks should examine whether a gain generalizes across tasks and modalities, not only whether it persists on the original task. Intensity here is inverted relative to the other eight response types: the relevant quantity is how quickly support is withdrawn, and fading itself is the

response being calibrated. Because every other response type in this taxonomy depends on this category to avoid becoming a permanent crutch, fading and transfer checks are treated in the implementation pathway in Section 7 as a standing requirement rather than an optional ninth item to be adopted only if resources allow.

The nine response types are not a menu to be worked through in order; they are organized around three underlying mechanisms that recur across the theoretical review in Section 2. Input, output, and feedback responses work by changing what linguistic evidence a learner receives or produces and how attention is directed within it. Scaffolded production and strategy instruction work by building capacity, mediated performance in the first case, transferable self-regulation in the second, that a learner can eventually exercise unassisted. Affective regulation, technology-mediated practice, peer mediation, and fading and transfer work chiefly by shaping the conditions under which the first two mechanism families can operate at all, whether by lowering anxiety, increasing practice volume, adding interactional variety, or managing the withdrawal of support. A response-selection decision that draws from only one of these three families without considering whether a condition-shaping response is also needed risks treating a conditions problem as though it were a capacity or evidence problem, which is precisely the kind of mismatch Proposition 6 in Section 6 is designed to detect.

Table 3. Adaptive response matrix: nine response types

Response type	Representative techniques	Target diagnostic signal	Fading criterion
Input modification	Calibrated simplification, elaboration, paced delivery (Krashen, 1985)	Comprehension gaps with intact decoding	Narrowing gap between input level and learner level
Output and interaction	Structured production under time pressure, negotiation of meaning	Fluency and grammatical control gaps visible mainly in production	Increased task complexity with reduced planning time
Corrective feedback and noticing	Recasts, explicit correction, attention-directing prompts (Schmidt, 1990)	Recurrent grammatical, pronunciation, or accuracy errors	Shift from explicit correction to learner self-monitoring
Scaffolded production	Graduated mediation hierarchy, dynamic assessment prompts (Vygotsky, 1978)	Large gap between assisted and unassisted performance	Movement up the mediation hierarchy toward independence
Strategy instruction	Modeled and practiced compensation and monitoring strategies (Oxford, 2017)	Strategic competence and metacognitive control gaps	Shift from teacher-modeled to learner-initiated strategy use

Affective regulation	Structured low-stakes practice, anxiety-reduction protocols (Gardner, 1985)	Language anxiety, low self-efficacy or willingness to communicate	Increased stakes and audience as anxiety declines
Technology-mediated adaptive practice	Computer-adaptive sequencing and difficulty stepping	Practice-volume constraints across domains	Automated reduction in practice frequency as accuracy stabilizes
Peer-mediated and collaborative tasks	Structured, rotating collaborative oral or written tasks (Rafi & Pourdana, 2023)	Interactional competence and willingness to communicate gaps	Increasing task openness as interaction stabilizes
Fading and transfer checks	Scheduled withdrawal of support with cross-task and cross-modality retesting	Any signal previously targeted by another response type	Retention and transfer confirmed at defined follow-up intervals

6 Testable propositions

Each proposition below follows the matched-versus-unmatched logic argued for in Section 4: the comparison of interest is a response matched to a specific diagnostic signal against an unmatched but otherwise similar response, not a response against no response at all, because only the matched-versus-unmatched comparison can isolate a matching advantage from a general effect of increased instructional attention. Each proposition also specifies the classification that must be made before treatment assignment, since a study that classifies learners after seeing outcomes cannot test a matching hypothesis without risking circularity.

Proposition 1. Learners whose comprehension gap is diagnosed with intact decoding will show larger gains from a calibrated input-modification response than from an uncalibrated, grade-level-default version of the same materials, provided the calibration tracks the specific dimension implicated by the diagnosis. A valid test should hold input dimension constant across conditions except the one calibrated, preregister the comparison, and report results separately by initial gap size. A null result would suggest that calibration adds little beyond generally simplified input, contrary to the dimension-specific logic this proposition assumes.

Proposition 2. Increasing task complexity and reducing planning time in output-and-interaction responses will accelerate fluency gains for learners whose fluency gap co-occurs with adequate grammatical knowledge, but will not accelerate, and may impede, gains for learners whose fluency gap co-occurs with a genuine knowledge gap. A valid test should classify learners on both dimensions before assignment, preregister the interaction, and report results separately by classification. Failure to find this interaction would argue against treating fluency-focused and knowledge-focused barriers as requiring different response intensity profiles.

Proposition 3. Corrective feedback delivered through an explicit attention-directing device will produce larger accuracy gains than the same corrective content delivered without an attention-directing device, consistent with the noticing hypothesis. A valid test should hold feedback content constant across conditions and vary only the attention-directing device, preregister the comparison, and track whether the effect persists after the device is withdrawn. A null result

would weaken the case for treating noticing as the operative mechanism behind this response type's effectiveness.

Proposition 4. Scaffolded production responses delivered at an intensity matched to the assisted-unassisted performance gap, rather than at a fixed intensity, will produce faster narrowing of that gap over repeated cycles. A valid test should randomize or counterbalance matched versus fixed intensity assignment, preregister the rate-of-narrowing comparison, and report delivery fidelity alongside outcome. A finding that fixed intensity narrows the gap as quickly as matched intensity would undercut the case for individualized dosage calibration central to this response type. Because scaffolded production is among the most resource-intensive response types to deliver with fidelity, this proposition should be tested alongside a cost accounting of mediator time per learner, so a matching advantage that requires impractical staffing ratios can be distinguished from one that is genuinely scalable.

Proposition 5. Strategy-instruction responses will show a delayed advantage over other response types, visible only after a period of independent practice rather than immediately post-instruction, because the response builds a transferable capacity rather than directly supplying linguistic knowledge. A valid test should measure outcomes at an immediate post-response point and again after a defined delay, preregister the predicted delayed-advantage pattern, and compare against a response type without this predicted delay. Absence of a delayed advantage would suggest the transfer-of-capacity mechanism this response type assumes is not operating as theorized.

Proposition 6. Reducing measured language anxiety through an affective-regulation response will increase the correspondence between a learner's classroom-observed performance and low-stakes diagnostic performance, with the effect concentrated among learners with high baseline anxiety. A valid test should measure anxiety, classroom performance, and low-stakes performance before and after the response, preregister the baseline-anxiety interaction, and report results by baseline-anxiety band. A uniform effect regardless of baseline anxiety would suggest the response operates through a mechanism other than anxiety reduction specifically.

Proposition 7. Technology-mediated adaptive practice will outperform teacher-paced practice on volume-constrained domains specifically, such as vocabulary breadth, but will not outperform teacher-mediated response on domains requiring qualitative judgment, such as pragmatic competence. A valid test should assign learners to matched domains under both delivery modes, preregister the domain-type interaction, and report results by domain. A finding that technology-mediated practice performs equally well across both domain types would argue against reserving this response type for volume-constrained domains as Section 5 proposes.

Proposition 8. Peer-mediated collaborative tasks will produce larger interactional-competence gains than equivalent individual tasks for learners with moderate willingness to communicate, but smaller gains for learners with very low willingness to communicate, whose peer interaction may be too anxiety-provoking to be productive without prior affective-regulation response. A valid test should classify learners by baseline willingness to communicate, preregister the interaction, and report results by classification. A uniform effect across willingness-to-

communicate levels would suggest peer mediation can be deployed without regard to this prior classification.

Proposition 9. A response that follows a predetermined fading rule will show better retention at follow-up than an equally effective response withdrawn abruptly or never withdrawn at all, because gradual fading allows the learner's own unassisted performance to be repeatedly tested and reinforced during withdrawal. A valid test should compare gradual, abrupt, and no-fading conditions holding initial response intensity constant, preregister the follow-up retention comparison, and report cross-task and cross-modality transfer, not only same-task retention. A finding that abrupt withdrawal retains as well as gradual fading would simplify the model by removing the need for a graduated fading rule.

Proposition 10. A response drawn only from the evidence-and-capacity-building response types, input modification, output and interaction, corrective feedback, scaffolded production, and strategy instruction, will underperform a response that also includes a condition-shaping type, affective regulation, technology-mediated practice, peer mediation, or fading and transfer, whenever the diagnosed signal co-occurs with a condition-level constraint such as high anxiety, low practice volume, or limited interactional variety. A valid test should classify learners on both the linguistic signal and the co-occurring condition before assignment, preregister the interaction, and report results separately by condition status. A null interaction would suggest that condition-shaping responses add little beyond evidence-and-capacity-building responses even when a condition-level constraint is present, weakening the three-family organization proposed in Section 5.

7 Implementation pathways

7.1 Response-menu definition

Response-menu definition requires a program to select, from the nine response types in Section 5, the subset its teachers can deliver with fidelity given their training and class sizes, and to write down, for each selected response, its target signal, intensity-calibration rule, and fading criterion before any learner receives it. A program that adopts every response type without regard to delivery capacity risks delivery-fidelity failure of the kind Mehri and Amerian (2015) document for resource-intensive mediation. The written protocol produced at this stage should state explicitly which response types the program is deliberately excluding for now, and why, so a later audit of practice can distinguish a considered omission from an unnoticed gap.

7.2 Teacher readiness

Teacher readiness assessment checks whether teachers can reliably classify a diagnostic signal at the signal-interpretation stage, deliver the selected responses at their intended intensity, and apply the stated fading rule rather than defaulting to either constant support or premature withdrawal. Training should include structured practice classifying illustrative, non-identifying signal profiles and calibrating intensity against a shared rubric before teachers apply the model to actual learners.

7.3 Pilot

The pilot phase applies the response-selection and delivery stages to a single class, with reassessment but without yet comparing matched against unmatched response assignment, so

response delivery and fading can be observed under real classroom conditions before any comparative claim depends on the results. Pilot data should test whether teachers can distinguish the nine response types in practice or tend to collapse several into a single generic differentiation strategy, a risk the taxonomy's explicit fading criterion are designed to reduce.

7.4 Protected evaluation

Protected evaluation compares matched against unmatched response assignment in a held-out class or cohort, with preregistered analysis testing the propositions in Section 6. This phase should report delivery fidelity alongside outcome, since a matching hypothesis cannot be fairly tested if the matched response was delivered with lower fidelity than the comparison response.

7.5 Advisory use

Advisory use permits teachers to receive a suggested response and intensity alongside their own classroom judgment, with a documented process for adjusting or overriding the suggestion and for recording why, so that override patterns can later be examined for evidence that the model's matching rules need revision rather than treating every override as teacher error. A model whose suggestions are overridden at a consistently high rate for a particular response type is providing evidence about that response type's real-world fit, and that evidence should feed back into the taxonomy rather than being treated only as a training gap to be corrected.

7.6 Classroom-wide adoption

Classroom-wide adoption extends the model across all classes in a program only after protected evaluation shows a matching advantage large enough to justify the training and delivery cost documented in the readiness phase; a model that shows a statistically detectable but practically small matching advantage may not justify full adoption once delivery cost is weighed against the alternative of investing that cost in the diagnostic identification work addressed elsewhere in this research program.

7.7 Monitoring

Ongoing monitoring tracks fading fidelity specifically, since a program under time pressure tends to drift toward either premature withdrawal of scaffolding or indefinite continuation of it, both of which defeat the acceleration this model is designed to produce, and monitoring should flag any response type whose observed fading pattern departs systematically from its stated criterion for review. Monitoring reports should be reviewed on a fixed schedule rather than only when an outcome disappoints, since a response type can drift toward indefinite continuation gradually and without any single disappointing outcome to trigger a review.

8 Discussion

This paper's central claim is that individualized instruction has a second problem beyond diagnosis: even an accurate, component-specific diagnostic signal does not by itself specify which instructional response fits it, at what intensity, or when that response should stop. The four-stage adaptive response cycle and nine-type taxonomy developed here are an attempt to make that second problem tractable and falsifiable rather than left to individual teacher intuition applied inconsistently across a program.

The most direct practical implication concerns dosage. Programs that adopt a response type, such

as scaffolded production or corrective feedback, without specifying an intensity-calibration rule are likely to deliver it inconsistently across teachers and classes, which will show up in evaluation as a weak or null average effect that actually conceals a real effect diluted by delivery variation. The delivery-fidelity requirement running throughout Sections 3 through 7 exists specifically to prevent this explanation from going undetected.

A second practical implication concerns training investment. The response-selection cross-check against a learner's broader aptitude profile, and the intensity-calibration logic tied to the assisted-unassisted performance gap, both require teacher judgment that goes beyond simply following a lookup table from signal to technique. Programs that underinvest in the readiness phase specified in Section 7 should expect the model to collapse in practice into a small number of generically applied techniques, which would reproduce the undifferentiated response problem this paper set out to solve. Methodologically, the matched-versus-unmatched comparison design required in Section 4 is the paper's central methodological commitment. A response-versus-no-response design cannot distinguish a matching advantage from a general effect of increased instructional attention, and several of the studies synthesized in Section 2 report exactly this kind of undifferentiated comparison; the propositions in Section 6 are written to require the stronger design. Methodologically, the requirement that intensity be reported as an actually-delivered quantity, not an intended one, has a further consequence worth stating explicitly: a program that cannot measure what intensity was actually delivered cannot participate in testing the propositions in Section 6 at all, regardless of how carefully it selects and delivers responses in practice. Investing in basic delivery-logging infrastructure is therefore a precondition for evaluation, not an optional refinement layered on top of it.

Ethically, the fading requirement is not only a pedagogical design choice but a safeguard against a specific failure mode: scaffolding that never withdraws can become indistinguishable from a permanently lowered expectation for a given learner, which is a version of the same labeling risk raised elsewhere in this research program with respect to diagnostic classification. A response type without a stated fading criterion should not be considered fully specified.

Who is positioned to deliver this model also matters. Macaro (2003) and Brown (2013) both argue that classroom teachers, not only specialist assessment staff, can generate and act on fine-grained diagnostic and response evidence once the instruments and decision rules are specified clearly enough to teach, and Heriansyah, Darni, Fajri, and Sahardin (2021) show that curriculum and assessment design shape how much of a given barrier is instructional in origin rather than learner-internal, which argues for training general classroom teachers in the signal-interpretation and response-selection stages rather than routing every case to a specialist. How this training should differ between a single secondary-school classroom and a multi-section postsecondary program, where staffing ratios, course structures, and student mobility differ substantially, is addressed in a later stage of this research program rather than here.

The model's main limitation mirrors its central claim: it specifies matching, calibration, and fading logic drawn from mechanisms each individually supported in the literature, but whether the full nine-type taxonomy, applied jointly through the four-stage cycle, produces acceleration beyond what any single well-delivered response type would produce alone is an empirical question the propositions in Section 6 pose rather than answer.

This paper assumes, rather than establishes, that a defensible diagnostic signal is available to feed the signal-interpretation stage; the taxonomy and evidentiary requirements for producing that signal are addressed elsewhere in this research program, and how a validated response

model of the kind proposed here should be trained, resourced, and monitored across the different institutional conditions of secondary and postsecondary settings is addressed in a later stage of that program as well. A further limitation concerns the three-family organization introduced in the synthesis following Section 5.9: evidence-and-capacity-building responses, and condition-shaping responses. This organization is a reading of the theoretical review in Section 2 rather than an empirical finding, and Proposition 10 is the paper's attempt to make the organization itself testable rather than assumed. If future evidence shows that condition-shaping responses add little even when a condition-level constraint is present, the taxonomy's implicit claim that both families are usually needed together would need revision, and the model would simplify toward evidence-and-capacity-building responses as the primary driver of acceleration.

9 Conclusion

The Adaptive Proficiency Acceleration Framework addresses the half of individualized second language instruction that a diagnostic signal alone cannot settle: which response fits the diagnosis, how much of it a given learner needs, and when it should stop. Its contribution is a four-stage decision architecture, signal interpretation, response selection, scaffolded delivery, and reassessment with fading, together with a nine-type response taxonomy that specifies, for each type, a target signal, an intensity-calibration rule, and a fading criterion. The taxonomy groups these nine types into three underlying mechanism families, evidence-and-capacity-building responses that change what a learner receives, produces, or can do independently, and condition-shaping responses that govern whether the first family can operate at all, a distinction formalized as Proposition 10 and intended to keep response design from defaulting to whichever technique is most familiar rather than whichever mechanism the diagnosis actually calls for.

No intervention is reported here, and none of the propositions in Section 6 should be treated as findings; each specifies the design a future study would need to test whether matching, calibrating, and fading response according to this model, delivered with recorded fidelity and evaluated against a genuine alternative rather than a passive comparison group, outperforms undifferentiated instruction. Future work executing the model empirically must use the matched-versus-unmatched design specified in Section 4, report delivery fidelity alongside outcome, and track fading and transfer, not only immediate post-response gains.

None of the nine response types proposed here is novel in isolation; each is drawn from an established strand of second language instruction research. What this paper adds is the requirement that every response type be specified completely, with a target signal, an intensity rule, and a fading criterion, before it is treated as ready for classroom use, a completeness standard the reviewed literature applies unevenly across response types even though the theoretical grounds for applying it consistently are already available.

Its central proposition is that generic differentiation, applying varied activities without a stated mechanism, target signal, intensity rule, and fading criterion, is not the same thing as individualized instruction, and that acceleration in proficiency growth depends on treating response selection and calibration with the same rigor diagnostic identification requires. The framework is ready for the empirical testing this paper deliberately does not attempt.

Read alongside the diagnostic component taxonomy developed elsewhere in this research program, the argument of both efforts together is that individualized second language instruction has two distinct failure points, coarse diagnosis and unmatched or miscalibrated response, and that treating either one alone as the whole problem will leave individualization more nominal

than real. This paper has addressed the second failure point on the assumption that the first has been adequately handled; a program that has solved neither should expect this framework's propositions to be difficult to test cleanly until diagnostic signal quality is itself established.

A reader who takes away only one point from this paper should take away the distinction between a response menu and a response architecture. A menu lists technique; an architecture specifies, for every listed technique, which signal warrants it, how intensely it should be delivered, and when it should stop. Generic differentiation, in practice, is usually a menu without an architecture, which is why it is difficult to evaluate, difficult to fade responsibly, and easy to apply inconsistently across teachers and classes even when every individual technique on the menu is, in isolation, well supported by research.

Declarations

Ethics approval and consent to participate. No human or animal data were collected for this conceptual and methodological paper. Any future empirical study applying the response model proposed here must obtain approvals appropriate to its participants, including, where minors are involved, parental consent and institutional review, records, and jurisdiction.

Consent for publication. Not applicable.

Availability of data and materials. No new dataset was generated or analyzed. The response taxonomy, intensity matrix, and evaluation criteria are reported in full within the manuscript.

Competing interests. The authors declare no competing interests.

Funding. The authors received no specific funding for this work.

Authors' contributions. Raissa Enyonam Amele Amessouwoe led conceptualization of the adaptive response cycle, drafted the theoretical review, and coordinated the manuscript. Kpoka Messanh Isaac developed the response taxonomy and intensity matrix, and contributed the technology-mediated response analysis. Julius Aderogba Odewabi developed the testable propositions and the implementation pathway, and conducted reference and citation verification. All authors reviewed and approved the final manuscript.

Acknowledgments. None.

References

- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency*. Continuum.
- Bachman, L. F., and Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.
- Badrin Kholid, A. Rahman, & L. A. Irawan. (2024). Implementing Diagnostic Assessment in Designing Differentiated Learning for English Language Learning at the Junior High Schools. *Journal of Language and Literature Studies*, 4(2), 445-458. <https://doi.org/10.36312/jolls.v4i2.1934>
- Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Brown, A. (2013). Multicompetence and Second Language Assessment. *Language Assessment Quarterly*, 10(2), 219-235. <https://doi.org/10.1080/15434303.2013.769551>
- Butler, Y. G. (2023). Self-assessment in second language learning. *Language Teaching*, 57(1), 42-56. <https://doi.org/10.1017/s0261444822000489>
- Butler, Y. G. (2024). Self-assessment in second language learning. *Language Teaching*.
- Caplan, N. A. (2019). Learning through language: A response to Polio, Keeping the language in second language classes. *Journal of Second Language Writing*, 46, 100677. <https://doi.org/10.1016/j.jslw.2019.100677>
- Council of Europe. (2020). *Common European Framework of Reference for Languages Companion Volume*.
- de Bot, K. (2014). The effectiveness of early foreign language learning in the Netherlands. *Studies in Second Language Learning and Teaching*, 3, 409-418. <https://doi.org/10.14746/ssllt.2014.4.3.2>
- DeKeyser, R. M. (2019). Aptitude-treatment interaction in second language learning. *Journal of Second Language Studies*, 2(2), 165-168. <https://doi.org/10.1075/jsls.00007.int>
- Dornyei, Z. (2005). *The Psychology of the Language Learner*. Lawrence Erlbaum.
- Dunkel, P. A. (1999). Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning & Technology*, 2(2), 77-93. <https://doi.org/10.64152/10125/25044>
- Ellis, R. (2008). *The Study of Second Language Acquisition*. Oxford University Press.
- Figueiredo, S. A., Martins, M. A., & Fernandes da Silva, C. (2013). Education and Learning in Second Language Contexts: Language and Diagnostic Assessment of Diverse Learners in Portuguese Schools. *The International Journal of Diversity in Education*, 13(1), 31-40. <https://doi.org/10.18848/2327-0020/cgp/v13i01/40078>
- Forgor, A. T. (2021). The impact of adopting of assessment for learning and adaptive learning system on students' foreign language performance. *LAPLAGE EM REVISTA*, 7(1), 221-228. <https://doi.org/10.24115/s2446-6220202171707p.221-228>
- Foster-Cohen, S. H. (2004). Relevance Theory and second language learning/behaviour. *Second Language Research*, 20(3), 189-192. <https://doi.org/10.1191/0267658304sr236ed>
- Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.
- Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. *Language Learning*, 66(2), 419-447. <https://doi.org/10.1111/lang.12157>
- Gardner, R. C. (1985). *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. Edward Arnold.

- Gass, S. M. (1989). Language Universals and Second Language Acquisition. *Language Learning*, 39(4), 497-534. <https://doi.org/10.1111/j.1467-1770.1989.tb00901.x>
- Ghanbarpour, M. (2017). A Qualitative Meta-synthesis of Research on Dynamic Assessment of Second/Foreign Language Learning: Implications for Language Teachers. *Journal of Language Teaching and Research*, 8(4), 731. <https://doi.org/10.17507/jltr.0804.12>
- Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 32(3), 317-336. <https://doi.org/10.1177/0265532214564505>
- Hashim, H., Salam, S., Nurul, S., & Syafiatun, N. (2018). The Designing of Adaptive Self-Assessment Activities in Second Language Learning using Massive Open Online Courses (MOOCs). *International Journal of Advanced Computer Science and Applications*, 9(9). <https://doi.org/10.14569/ijacsa.2018.090937>
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heriansyah, H., Darni, P., Fajri, R., & Sahardin, R. (2021). Competency-based curriculum, relevant second language learning theories, and its language assessment. *English Education Journal*, 12(3), 416-427. <https://doi.org/10.24815/eej.v12i3.21631>
- In'nami, Y. (2013). Second-language vocabulary assessment research: Issues and challenges. *Vocabulary Learning and Instruction*, 2(1), 74-81. <https://doi.org/10.7820/vli.v02.1.innami>
- Isbell, D. R. (2021). Can the Test Support Student Learning? Validating the Use of a Second Language Pronunciation Diagnostic. *Language Assessment Quarterly*, 18(4), 331-356. <https://doi.org/10.1080/15434303.2021.1874382>
- Jakobovits, L. A. (1969). Second language learning and transfer theory: A theoretical assessment. *Language Learning*, 19(1-2), 55-56. <https://doi.org/10.1111/j.1467-1770.1969.tb00933.x>
- Johnson, J. S. (1992). Critical Period Effects in Second Language Acquisition: The Effect of Written versus Auditory Materials on the Assessment of Grammatical Competence. *Language Learning*, 42(2), 217-248. <https://doi.org/10.1111/j.1467-1770.1992.tb00708.x>
- Ketabi, S., & Ketabi, S. (2014). Classroom and Formative Assessment in Second/Foreign Language Teaching and Learning. *Theory and Practice in Language Studies*, 4(2). <https://doi.org/10.4304/tpls.4.2.435-440>
- Koda, K. (2007). Reading and Language Learning: Crosslinguistic Constraints on Second Language Reading Development. *Language Learning*, 57(s1), 1-44. <https://doi.org/10.1111/0023-8333.101997010-i1>
- Koizumi, R. (2015). Second language vocabulary assessment studies: Validity evidence and future directions. *Vocabulary Learning and Instruction*, 4(1), 36-46. <https://doi.org/10.7820/vli.v04.1.koizumi>
- Kostolanyova, K., & Simonova, I. (2020a). Adaptive e-learning model for learning English as a second/foreign language. *International Journal of Innovation and Learning*, 27(3), 274. <https://doi.org/10.1504/ijil.2020.10028282>
- Kostolanyova, K., & Simonova, I. (2020b). Adaptive e-learning model for learning English as a second/foreign language. *International Journal of Innovation and Learning*, 27(3), 274. <https://doi.org/10.1504/ijil.2020.106811>
- Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Longman.

- Lanteigne, B. (2017). Unscrambling jumbled sentences: An authentic task for English language assessment? *Studies in Second Language Learning and Teaching*, 7(2), 251-273. <https://doi.org/10.14746/ssllt.2017.7.2.5>
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50. <https://doi.org/10.1016/j.jslw.2012.11.003>
- Li, M. (2024). Leveraging ChatGPT for Second Language Writing Feedback and Assessment. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1-11. <https://doi.org/10.4018/ijcallt.360382>
- Macaro, E. (2003). Second language teachers as second language classroom researchers. *The Language Learning Journal*, 27(1), 43-51. <https://doi.org/10.1080/09571730385200071>
- Mehri, E., & Amerian, M. (2015). Challenges to Dynamic Assessment in Second Language Learning. *Theory and Practice in Language Studies*, 5(7), 1458. <https://doi.org/10.17507/tpls.0507.19>
- Menaka, G., & Sankar, G. (2019). The language learning assessment using technology for the second language learners. *International Journal of Linguistics, Literature and Culture*, 5(4), 1-6. <https://doi.org/10.21744/ijllc.v5n4.674>
- Mislevy, R. J., & Yin, C. (2009). If Language Is a Complex Adaptive System, What Is Language Assessment? *Language Learning*, 59(s1), 249-267. <https://doi.org/10.1111/j.1467-9922.2009.00543.x>
- Mohamadi, Z. (2017). Task engagement: a potential criterion for quality assessment of language learning tasks. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1). <https://doi.org/10.1186/s40862-017-0025-z>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Navracsics, J. (2014). Input or intimacy. *Studies in Second Language Learning and Teaching*, 3, 485-506. <https://doi.org/10.14746/ssllt.2014.4.3.6>
- Neumann, H. (2014). Teacher assessment of grammatical ability in second language academic writing: A case study. *Journal of Second Language Writing*, 24, 83-107. <https://doi.org/10.1016/j.jslw.2014.04.002>
- Nida, E. A. (1956). Motivation in second language learning. *Language Learning*, 7(3-4), 11-16. <https://doi.org/10.1111/j.1467-1770.1956.tb01198.x>
- Norris, J. (2001). Concerns with Computerized Adaptive Oral Proficiency Assessment. *Language Learning & Technology*, 5(2), 99-105. <https://doi.org/10.64152/10125/25131>
- Oh, S. (2019). Second Language Learners' Use of Writing Resources in Writing Assessment. *Language Assessment Quarterly*, 17(1), 60-84. <https://doi.org/10.1080/15434303.2019.1674854>
- Onwubuya, L. I., Fawehinmi, Y. O., Sunday, D., Igwenagu, M. O., Adeniran, A. O., & Agyapong, E. A. (2026). Adaptive learning systems powered by artificial intelligence for STEM education improvement. *Asian Journal of Education and Social Studies*, 52(6), 553-566. <https://doi.org/10.9734/ajess/2026/v52i61705>
- Oxford, R. L. (2017). *Teaching and Researching Language Learning Strategies*. Routledge.
- Poole, D. (1992). Language Socialization in the Second Language Classroom. *Language Learning*, 42(4), 593-616. <https://doi.org/10.1111/j.1467-1770.1992.tb01045.x>
- Pourdana, N. (2022). Impacts of computer-assisted diagnostic assessment on sustainability of L2 learners' collaborative writing improvement and their engagement modes. *Asian-Pacific*

- Journal of Second and Foreign Language Education, 7(1).
<https://doi.org/10.1186/s40862-022-00139-4>
- Pulido, D. (2003). Modeling the Role of Second Language Proficiency and Topic Familiarity in Second Language Incidental Vocabulary Acquisition Through Reading. *Language Learning*, 53(2), 233-284. <https://doi.org/10.1111/1467-9922.00217>
- Rafi, F., & Pourdana, N. (2023). E-diagnostic assessment of collaborative and individual oral tiered task performance in differentiated second language instruction framework. *Language Testing in Asia*, 13(1). <https://doi.org/10.1186/s40468-023-00223-7>
- Roever, C. (2007). DIF in the Assessment of Second Language Pragmatics. *Language Assessment Quarterly*, 4(2), 165-189. <https://doi.org/10.1080/15434300701375733>
- Safari, F., & Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *Journal of Second Language Writing*, 60, 101007. <https://doi.org/10.1016/j.jslw.2023.101007>
- Sang, F. (1983). Second language acquisition and second language learning. *System*, 11(1), 84-89. [https://doi.org/10.1016/0346-251x\(83\)90013-1](https://doi.org/10.1016/0346-251x(83)90013-1)
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Shin, J. A., & Christianson, K. (2011). Structural Priming and Second Language Learning. *Language Learning*, 62(3), 931-964. <https://doi.org/10.1111/j.1467-9922.2011.00657.x>
- Spinner, P. (2011). Second language assessment and morphosyntactic development. *Studies in Second Language Acquisition*, 33(4), 529-561. <https://doi.org/10.1017/s0272263111000301>
- Spolsky, B. (1969). Attitudinal aspects of second language learning. *Language Learning*, 19(3-4), 271-275. <https://doi.org/10.1111/j.1467-1770.1969.tb00468.x>
- Ströbel, M., Kerz, E., & Wiechmann, D. (2020). The Relationship Between First and Second Language Writing: Investigating the Effects of First Language Complexity on Second Language Complexity in Advanced Stages of Learning. *Language Learning*, 70(3), 732-767. <https://doi.org/10.1111/lang.12394>
- Sun, Z., Xu, P., & Wang, J. (2022). Dynamic Assessment of the Learning Potential of Chinese as a Second Language. *Language Assessment Quarterly*, 20(1), 127-142. <https://doi.org/10.1080/15434303.2022.2151911>
- Teló, C., Kivistö de Souza, H., O'Brien, M. G., & Carlet, A. (2024). Second Language Sentence Stress Assignment: Self- and Other-Assessment. *Language Learning*, 75(3), 832-862. <https://doi.org/10.1111/lang.12682>
- Thomas, M. (1994). Assessment of L2 Proficiency in Second Language Acquisition Research. *Language Learning*, 44(2), 307-336. <https://doi.org/10.1111/j.1467-1770.1994.tb01104.x>
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- Wakabayashi, S. (2009). Lexical learning in second language acquisition: optionality in the numeration. *Second Language Research*, 25(2), 335-341. <https://doi.org/10.1177/0267658308100293>
- Wanek-Klimczak, E. (2018). Review of Second language pronunciation assessment: Interdisciplinary perspectives (T. Isaacs & P. Trofimovich, Eds.). *Studies in Second Language Learning and Teaching*, 8(3), 701-706. <https://doi.org/10.14746/ssllt.2018.8.3.9>

- Xu, H., & Li, G. (2024). CEFR-informed learning, teaching and assessment: A practical guide. *Studies in Second Language Learning and Teaching*, 14(2), 365-370. <https://doi.org/10.14746/ssllt.32669>
- Yang, S. (2021). Constructing Diagnostic Reading Assessment Instruments for Low-level Chinese as Second Language Learners. *Reading in a Foreign Language*, 33(2), 212-237. <https://doi.org/10.64152/10125/67401>
- Yeotak Yoon, 김정우, & Min, H. (2008). A Study of the Diagnostic Assessment of Basic Competency in Middle School Korean Language Learning. *Korean Language Education Research*, 32, 305-340. <https://doi.org/10.20880/kler.2008..32.305>
- Zhang, L. J. (2013). Second language writing as and for second language learning. *Journal of Second Language Writing*, 22(4), 446-447. <https://doi.org/10.1016/j.jslw.2013.08.010>
- 이연숙. (2014). Assessing Second Language Grammar Using Cognitive Diagnostic Assessment Method. *The Journal of Studies in Language*, 30(2), 395-415. <https://doi.org/10.18627/jslg.30.2.201408.395>