World Journal of Education and Humanities
ISSN 2687-6760 (Print) ISSN 2687-6779 (Online)
Vol. 8 No. 4, 2026
www.scholink.org/ojs/index.php/wjeh

Original Paper

AI-Enabled Production-Oriented Approach: Conceptual Development and a Research Framework A Systematic Review of 46 Studies with Reading as a Test Case

Xuan Sun[1*]

[1] Kunming College of Arts and Sciences, Kunming 650222, Yunnan, China

[*] Corresponding author. E-mail: alexandriaxsun@163.com

Received: October 18, 2026 Accepted: November 09, 2026 Online Published: November 27, 2026

doi:10.22158/wjeh.v8n4p109 URL: http://dx.doi.org/10.22158/wjeh.v8n4p109

Abstract

Teachers and curriculum writers in Chinese foreign language education have used the label AI-enabled production-oriented approach (AI-POA) for several years, but it has no agreed definition, dimensions or boundary, and studies that appear to address the same arrangement cannot be compared. This paper treats the difficulty as conceptual. A systematic search of 1,373 records from OpenAlex and four from a targeted Chinese search yielded 46 studies published between 2017 and 2026, coded for design, technology role, POA stage and outcome. The corpus is deliberately mixed and is not a set of 46 instances: eight are POA sources written before generative tools existed, twenty-six combine POA with AI, ten examine AI in second language reading without POA, and two are reviews. Four roles for the technology are abduced—input source (n = 11), interlocutor (n = 8), evaluator (n = 10) and design partner (n = 7)—and the distribution of abilities is the more telling result: of the twenty-six POA-with-AI studies eleven address writing and one reading, while nine of the ten reading studies make no use of POA. The paper then specifies a definition with four necessary constituents, a 3×3 grid of affordance by stage whose sixty-one coded activations fall unevenly (thirty-three generative, thirteen dialogic), and a three-layer framework linking technology, pedagogy and ability through relations of fit. Four propositions follow, two of which structural equation modelling can test now; reading is treated as the test case.

Keywords

AI-enabled POA, production-oriented approach, generative AI, concept building, systematic review, second language reading

1. Introduction

1.1 A Label in Search of a Concept

The phrase *rengong zhineng chandao* (人工智能产导) circulates in Chinese conference papers, teaching competitions and curriculum documents, alongside 人工智能赋能产出导向法, AI 赋能 POA and 智能 POA, with no indication that the names refer to different things. Practitioners need a short label for an arrangement they already run, in which generative tools supply material, mediate dialogue and mark output inside a POA cycle; researchers need a definition that can be operationalised, dimensions that can be measured, and a boundary that excludes what the definition does not cover.

A term becomes a concept when it can be defined, decomposed and embedded. The label fails on all three and has no agreed definition, which is weaker than a dispute. The cost falls on the literature, where one study adds a text-generation step to a conventional unit and another rebuilds assessment around machine scoring; both claim to examine AI-supported POA, their findings are not comparable, and reviews pool them as though they were.

1.2 Why Conceptual Work is Needed now

Three developments make the absence costly. Technical: multi-turn models made material generation, interactive scaffolding and evaluative feedback available in one system, roughly the supports POA has always required (Wen, 2015; 2020). A method designed around what teachers could not afford now meets a technology that can afford part of it, and whether the potential is realised depends on design, which needs a vocabulary.

Scholarly: in 2025 *Modern Foreign Languages* devoted a section to AI-enabled POA (Wen, 2025a; Zhang, 2025; Qiu, 2025), a fourth paper on interpreter training being excluded here on grounds of ability rather than quality (Sun & Qi, 2025). It is programmatic rather than cumulative, each paper describing a practice and none proposing a construct that another could test (see also Zhu, 2026).

Policy: three documents commit the sector to AI-supported teaching and set the frame for classroom assessment—the *Outline of the Plan for Building a Strong Education Nation (2024-2035)*, the *Overall Plan for Deepening the Reform of Educational Evaluation in the New Era*, and the *Opinions on Accelerating the Digitalisation of Education*.

1.3 Questions, Scope and Approach

Three questions organise the review: how do existing studies describe the combination of AI and POA, and where do their accounts converge and diverge (RQ1); through what stages has the combination evolved (RQ2); and can a construct be specified that is operational enough to code and testable enough to fail (RQ3)?

Two scope decisions govern every table. The construct is general, because POA is a skill-neutral architecture whose three stages describe a teaching cycle rather than a language modality; a reading-specific definition would inherit the architecture, misplace the boundary and make the writing literature, where POA is most fully implemented, inadmissible. Reading is nonetheless the test case, because the evidence there is thinnest relative to the claim and measurement is hardest (Section 6). The

method is a concept-building rather than an effectiveness review, the latter needing homogeneous treatments that this literature does not supply.

2. Method

2.1 Design and the Architecture of the Corpus

The review follows PRISMA 2020 (Page et al., 2021) with guidance from Xiao and Watson (2019). AI-POA is not a controlled term in any thesaurus, so a literal search returns nothing; the review therefore searches the POA cluster and the generative-AI cluster separately and maps their intersection, which is where the construct should be.

The corpus is heterogeneous, and the heterogeneity carries information. *POA theory sources* (n = 8) state the three stages, principles and hypotheses, contain no AI, and anchor the construct in POA's own specification. *Core studies* (n = 26) combine POA with AI in a design, implementation or evaluation and are the only studies showing what the combination looks like in practice. *Contextual studies* (n = 10) examine AI in second language reading without POA and supply the outcome evidence the core studies lack. Two *reviews* are maps rather than evidence. Strata are reported separately in every table because the claim that 46 studies are 46 instances cannot stand; what stands is that 46 studies are the evidence base from which the construct is abduced.

2.2 Search Strategy

OpenAlex was the international source: it covers Crossref and much of the Scopus and Web of Science journal list, which made it a workable substitute for two subscription databases that were unavailable. CNKI, restricted to CSSCI and Peking University core journals, was the Chinese source; because it does not permit record export for this string, the search was screened in the browser and verified against publisher records, so only four items are counted.

Four strings were applied to title and abstract for 1 January 2015 to 31 October 2026 and are given in full in Appendix A. S1 was run in two forms: the full name returned 602 records, the acronym qualified by a language-teaching term returned 237, of which 185 duplicated the first set and 52 did not, 47 of the 52 applying POA to a course with no technology and five describing an AI component too thinly to code; that supplement ran after screening and no included study comes from it. S3 now includes "large language model" alongside the model names, in line with the Chinese string. S2 is restricted to reading by design, since POA studies name POA in the title or abstract and enter through S1 (Section 8). The four strings yielded 1,373 records plus four Chinese items, 1,338 after de-duplication.

2.3 Inclusion and Exclusion

Inclusion was stratified. A record qualified if it stated or revised POA theory; combined POA with a generative or AI-based tool; examined AI in second language reading or writing and measured a language outcome; or reviewed the relevant literature systematically — in each case concerning foreign language teaching, peer-reviewed, and published between 2015 and 2026. Exclusion removed technical computer-science papers with no instructional setting, studies outside foreign language education,

pre-generative technology making no claim about AI, records without accessible full text, and duplicates.

The third exclusion needs a word, since an earlier version appeared to contradict itself: seven included studies date from 2015 to 2020 and all seven are POA theory sources with no technology, untouched by a rule aimed at studies claiming an AI contribution while describing rule-based systems. Where the POA component was decorative, the record was excluded, and that judgement is the main source of discretion in the review.

2.4 Screening

Screening proceeded in three passes. Title and keyword screening removed 1,186 records, most because the topic was not foreign language teaching or the acronym had matched something else. Abstract and full-text assessment removed 100, largely technical studies, non-empirical commentaries and studies where reading or POA appeared only in passing. Quality appraisal removed 6. The final set is 46 studies, 42 in English and 4 in Chinese. Figure 1 reports the flow, with the Chinese branch shown at the point where it entered.

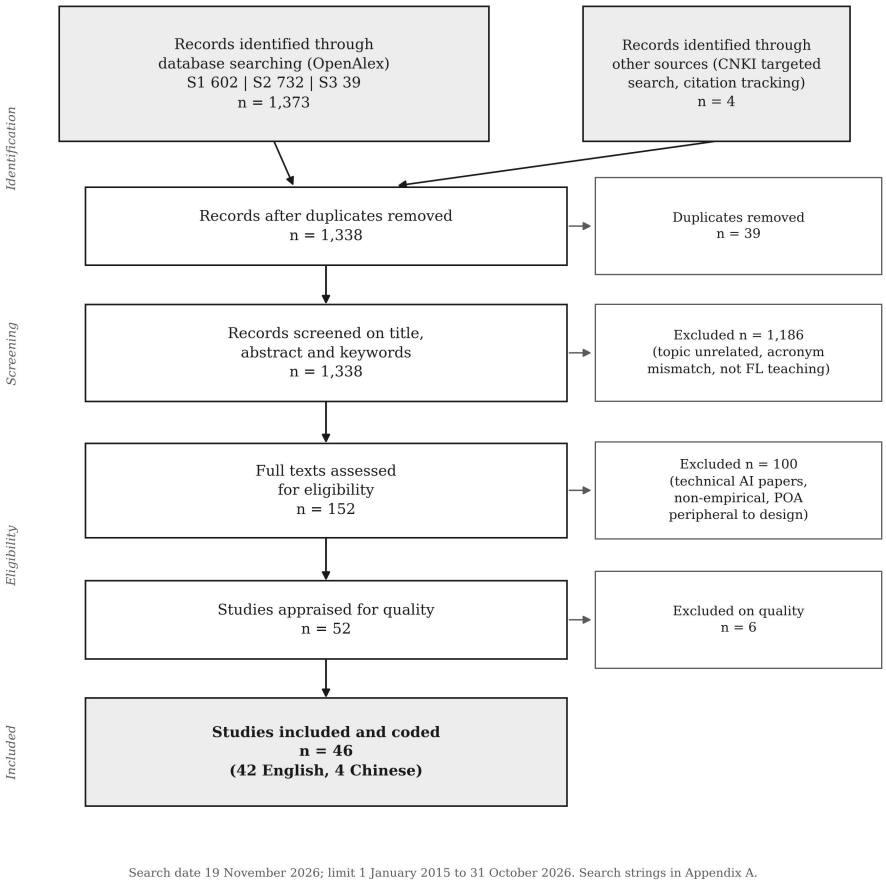


Figure 1. Flow of Studies through Identification, Screening and Inclusion

2.5 Coding and Appraisal

Seven dimensions were coded—year and setting, stratum, design, technology role, POA stage, ability, direction of effect—with definitions in Appendix B and study-level codes in Appendix C.

Coding was carried out by the author alone. Agreement over time was checked by recoding ten randomly sampled studies (21.7%) after two weeks, giving Cohen's kappa of .86. This is intra-coder agreement, not inter-coder reliability; no second coder was available, and Section 8 states what follows.

Quality was appraised by method type: empirical studies with a shortened Mixed Methods Appraisal Tool (Hong et al., 2018), conceptual papers with the JBI checklist for text and opinion (McArthur et al., 2020). Fifty-two records were appraised, six rated low and excluded; the 46 retained split into twelve high and thirty-four medium, and none met all four MMAT criteria, largely because implementation fidelity was rarely described.

2.6 Synthesis

Synthesis followed Thomas and Harden's (2008) three stages—free coding, descriptive themes, analytical themes—with Braun and Clarke's (2006) procedure used within each study to reduce the account to a triple of what the AI did, which stage it entered and what resulted. Braun and Clarke are cited for the procedure rather than the synthesis method, since their framework was built for single datasets. The third step is interpretive and no claim of reproducibility is made for it; the coding frame and the 46 triples are printed in Appendices B and C.

2.7 How the Corpus Is Read

One sentence governs the rest of the paper: the 46 studies are the evidence from which the construct is abduced, not a set of cases that satisfy it. Applied literally, the definition in Section 4.1 is met by one study, and that is not a defect, since a concept-building review needs material showing the phenomenon in partial forms. The theory sources fix the stages, the core studies show what practitioners have built with them, the contextual studies show what the technology achieves without them, and the construct is the difference between the last two.

3. Findings

3.1 The Shape of the Field, and Three Periods

Seven studies appeared between 2015 and 2020, two between 2021 and 2023, and thirty-seven between 2024 and 2026 (six, sixteen, fifteen): four fifths of the literature is under three years old. Growth of that shape follows the release of conversational models at the end of 2022: the tool arrives, practice adapts, writing follows. The cost is that questions are inherited from the technology rather than derived from the pedagogy.

2017 to 2020: consolidation without AI. All eight theory sources date from this window, and the question was whether the three stages could be specified precisely enough to be taught and observed: statements on motivating (Zhang, 2020), enabling (Qiu, 2020) and the updated architecture (Sun, 2020), with commentary from Ellis (2017) and Pu (2017).

2021 to 2023: first hybrids. Two studies, both placing AI inside an existing unit rather than redesigning it (Wang & Yan, 2022; Luo & Liang, 2022). Neither proposes a construct; Luo and Liang (2022) report a gain, but without an effect size or a comparison beyond intact classes, so the pair establish feasibility rather than effectiveness.

2024 to 2026: the generative turn. Thirty-seven studies, twenty-four of them core. The question shifts twice: which stage benefits from the technology, then how a programme should be organised across a curriculum. Proliferation is real and so is fragmentation: the core studies cite one another rarely, and no two share an instrument.

3.2 Four Roles for the Technology

Thematic synthesis produced four clusters among the 36 studies that involve AI, distinguished by what the tool does inside the cycle (Table 2). The paragraphs below give the argument; Appendix C gives the codes.

Table 2. Composition of the Corpus by Cluster

Cluster	n	English	Chinese	Designs	Ability in focus	Contextual
AI as input source	11	11	0	D 5, Q 3, T 3	writing 6, reading 3, general 2	2
AI as interlocutor	8	7	1	S 3, Q 2, C 1, D 1, T 1	speaking 3, reading 2, writing 2, general 1	3
AI as evaluator	10	10	0	D 7, Q 2, S 1	reading 5, writing 3, general 2	5
AI as design partner	7	4	3	T 5, C 2	general 6, writing 1	0
POA theory sources	8	8	0	T 8	not applicable	0
Reviews	2	2	0	R 2	not applicable	0
Total	46	42	4	T 17, D 13, Q 7, S 4, C 3, R 2	reading 10, writing 12, speaking 3, general 11, none 10	10

Note. Designs: T theoretical, D design or development, Q quasi-experiment, S survey, C case study, R review. A study may be counted in more than one design column only where it reports two components; the totals here assign the dominant design.

AI as a source of input. 11 studies address the generation and adaptation of material, and the reported contribution is fit rather than volume: textbooks adapt to a class average, whereas generation allows adaptation to a group or to a student within one lesson (Wang & Yan, 2022; Lin et al., 2025; Ironsi &

Bostanci, 2026). The inverse procedure, in which the model attempts the production task first so that its vocabulary becomes the input target, appears in Zhang (2025), coded as design partner. The risk is that ease substitutes for judgement: without a specified production goal a teacher generates more material, not better material.

AI as an interlocutor. 8 studies, the most developed conceptually: a three-way arrangement of teacher, model and student in the enabling stage (Qiu, 2025); a writing unit rebuilt as five exchanges with successive drafts (Li, 2024); model-supported practice that gave readers room to ask questions, some more willing to expose difficulty than in class (El Hassan & Al Salwah, 2025); and three speaking studies treating the model as rehearsal partner (Dong, 2026; Liu & Liang, 2026; Wang et al., 2026). POA's output-driven hypothesis holds that production is motivated by communicative need, and a model supplies an interlocutor perceived as less judgemental than a teacher or peer. The risk is symmetrical: a partner that always understands partial language removes the pressure to repair it, and repair is where much of the learning sits.

AI as an evaluator. 10 studies, dominated by automated scoring and item generation and organised by a tension between efficiency and validity. Models mark generously, catch language errors more reliably than structural ones, generate items with weak distractors (Lin & Chen, 2024) and produce diagnostic tests skewed towards remembering and understanding (Khamrassamee et al., 2025). Wen's (2016) collaborative assessment acquires a new division of labour: the machine screens, the teacher attends to representative cases. Screening is defensible; adjudication is not.

AI as a design partner for teachers. 7 studies, the only cluster in which Chinese-language work outnumbers English by more than three to one — three of seven, a concentration but not a monopoly, and calling the cluster "mainly Chinese" overstated it. Wen (2025a) sets out two principles and a six-step workflow for textbook writing, Zhang (2025) supplies the account of textbook use such a workflow presupposes, and Wen (2025b) extends the argument to the teaching model rather than the unit. The English items are thinner, reporting a case study of textbook use (Wu, 2026) or proposing designs without reporting use (Guo, 2024; Ma, 2026; Zhao, 2026). The cluster locates the first effect where it falls: AI changes teacher work before student learning, and where design capacity does not rise the technology accelerates the production of mediocre units.

3.3 How Theory Is Used

Thirty-nine studies (84.8%) name a framework and seven none (Table 4). The pattern is a division of labour that has become a split: technology-side studies explain why a student accepts a tool but not what the tool should do instructionally, pedagogy-side studies explain how to design a unit but not why it works. Eleven name POA, most as a design template, and the one hypothesis it makes testable — that production drives learning through the enabling stage — is tested by none.

Table 4. Theoretical Frameworks Named

Framework	n	%
Technology acceptance model and variants	9	19.6
Self-determination theory	6	13.0
Sociocultural theory	5	10.9
POA	11	23.9
Other (cognitive load, TPACK, OBE)	8	17.4
None named	7	15.2
Total	46	100.0

Note. Each study is counted once, under the framework it treats as primary. Where two were named, the one used to generate hypotheses was coded.

3.4 Methods and the Strength of the Evidence

Table 5. Design by Ability in Focus

Design	Reading	Writing	Speaking	General	None	Total
Theoretical	0	2	0	7	8	17
Design or development	4	7	0	2	0	13
Case study	0	0	1	2	0	3
Quasi-experiment	3	3	1	0	0	7
Survey	3	0	1	0	0	4
Review	0	0	0	0	2	2
Total	10	12	3	11	10	46

Note. “General” covers studies that address college English or foreign language teaching without naming a single ability.

The distribution supports a modest claim. Seventeen studies (37.0 per cent) are theoretical or programmatic, thirteen (28.3%) design proposals, seven quasi-experiments and four surveys. There is no randomised trial, and the seven quasi-experiments are the only comparative outcome evidence — enough to say that AI-POA is feasible in particular settings, not that it works in general. They share three weaknesses: interventions of one or two units, no delayed post-test, and no report of implementation fidelity; where implementation is undescribed, neither a null nor a positive result can be attributed to the design. Of fifteen studies reporting a positive result, six are quasi-experiments and two surveys of perceived benefit; the rest describe an implementation and present it as a gain.

3.5 Five Gaps

The concept is missing. No study uses AI-POA as a named construct, and the labels in circulation are not distinguished, so studies cannot cite one another's constructs.

The mechanism is a black box. Most studies report that something worked; few test a path. Whether the effect runs through motivation to produce, through the quality of input or through feedback cannot be settled.

Boundary conditions are absent. Self-regulation, prior attainment and AI literacy are plausible moderators and remain untested.

Reading and POA do not meet. Ten studies measure reading, twelve address writing, and one sits in both sets: reading is studied without POA and POA without reading. Comprehension is not observable, so a gain is hard to distinguish from a substitution.

Negative findings are scarce. One study reports a null result and two report mixed results, none treating the null as its primary finding; offloading, shallow processing and dependence are acknowledged as risks rather than studied as phenomena.

4. Defining the Construct

4.1 The Definition

The review supports the following.

AI-enabled production-oriented approach (AI-POA) is a form of human–machine collaborative teaching in which the motivating, enabling and assessing cycle of POA supplies the pedagogical core, three affordances of generative AI (generative, dialogic, evaluative) supply the mediating tools, and the teacher sets goals, vets material, and arbitrates machine output. It is a variant of POA rather than a separate method.

Two features of the wording carry weight. "Variant" places the construct inside an existing theory, a weaker claim than novelty and a more useful one, since it inherits POA's scope conditions along with its architecture. "Arbitrates" places the teacher inside the arrangement rather than outside it.

4.2 Four Constituents

The definition has four parts, and a case that lacks any of them is not an instance.

The pedagogical core is unchanged. Learning-centredness, the integration of learning and use, and output-first sequencing remain as POA specifies them. An arrangement that begins with a model essay to be imitated is not AI-POA, because the driving stage has been removed.

The affordances cover the cycle. The requirement is not that a single tool exercise all three affordances, nor that the technology be present at every stage, but that each stage is supported by at least one affordance and all three are represented across the cycle. A design using generation in the enabling stage and evaluation in the assessing stage satisfies it; one that generates worksheets and nothing else does not, and is better called AI-assisted POA.

The teacher sets goals, vets material and arbitrates output. Students using a model on their own initiative do not constitute AI-POA; what is missing is the three functions only a teacher can perform, namely specifying the production goal, vetting what the machine supplies, and adjudicating between a machine score and a student's defence of an answer.

The relation is collaborative rather than substitutive. The machine takes high-frequency, standardisable work and the teacher takes judgement, value and relationship; substitution implies the teacher's withdrawal, at which point the construct fails on its own terms.

The second constituent is the strictest: met in full by one study (Wang & Yan, 2022) and partially by perhaps a dozen. *AI-POA* names the construct as defined; *AI-assisted POA* names arrangements meeting some but not all of the constituents, which is most of the corpus; *AI-supported language teaching* is the umbrella covering any use of a model in a classroom, including use the teacher neither knows about nor sanctions.

4.3 A Grid of Affordance by Stage

A definition that cannot be decomposed cannot be measured. The proposal is a 3×3 grid, with the three affordances as rows and the three POA stages as columns (Table 1); each cell names a practice that can be designed, observed and coded.

Table 1. A 3×3 Grid of Affordance by Stage

Affordance / stage	Motivating	Enabling	Assessing
Generative	Scenario and task generation	Tiered and multimodal materials	Sample scripts and rubric drafting
Dialogic	Gap diagnosis through questioning	Just-in-time glossing at the point of breakdown	Interrogating a score; lodging an appeal
Evaluative	Baseline diagnosis	Process judgement of comprehension	Scoring with revision advice

The cells are not equally occupied, and the unevenness is the finding (Table 3).

Table 3. Studies Activating each cell of the Grid

Affordance / stage	Motivating	Enabling	Assessing	Total
Generative	10	16	7	33
Dialogic	1	8	4	13
Evaluative	2	5	8	15
Total	13	29	19	61

Note. Cell entries are studies. A study that places the technology in two stages contributes to two cells, so the 36 AI-involving studies yield 61 activations.

The generative row and the assessing column are crowded: 33 of 61 activations are generative and 19 fall in the assessing stage. The dialogic row is thin at 13 and one of its cells is nearly empty — a single study places dialogue in the motivating stage, against ten that place generation there. The emptiness fits the standing criticism of AI-supported reading: where dialogue is not designed in, the student opens a translation app at the point of difficulty and returns to the text unchanged.

4.4 Adjacent Constructs

Four neighbouring terms are easily confused with this one. *Computer-assisted language learning* places technology around a syllabus the teacher owns, whereas AI-POA places it inside the cycle; the test is what happens when the technology is removed, since a CALL arrangement continues and an AI-POA unit collapses at the assessment stage. *Automated writing evaluation* takes the product as its object and can stand alone, while AI-POA takes the cycle as its object and cannot. *Generative AI-assisted language learning* covers any student use of a model, including unsanctioned use, of which AI-POA is the institutionalised subset.

4.5 Conceptual Boundaries

Two boundaries define what the construct covers, and both are properties of the construct rather than of the evidence. The pedagogical boundary excludes orientations that do not centre on production: appreciative reading of literature, where the point is receptivity, and examination coaching, where the point is strategy, are not instances. The technical boundary requires a model capable of multi-turn exchange and extended text, which excludes rule-based and retrieval-based systems because they cannot sustain the dialogic affordance on which the second constituent depends. This boundary moves with the technology, the weakness of any construct defined against a moving line.

4.6 Evidence Boundaries and Application Conditions

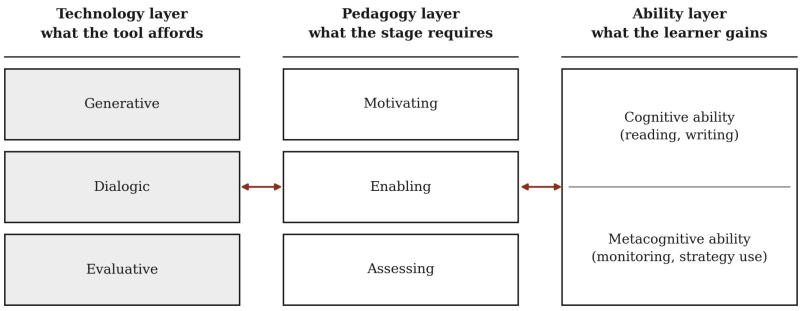
Two further limits are not conceptual, and separating them from Section 4.5 prevents a category error: a construct is not refuted by the fact that its evidence comes from one setting. The contextual boundary is a property of the evidence. The corpus comes from large classes with constrained teacher capacity, mostly in Chinese universities, with a minority of studies from Vietnam, Thailand, Turkey and Iran; where the teacher can already supply individual response, the marginal value of machine feedback falls. The framework in Section 5 should therefore be read as a claim about large-class tertiary instruction.

The ethical boundary is a condition of application. Student interaction data require consent, a retention period and a defined secondary use, and generated material has unsettled copyright status.

5. A Framework of Affordance, Stage and Ability

5.1 Three Layers

The framework has three layers, each with its own unit of analysis (Figure 2).



Each affordance is crossed with each stage, giving nine cells (Tables 1 and 3).

The reading instantiation of the ability layer is set out in Section 6.

Figure 2. A Three-layer Framework of Affordance, Stage and Ability. Both Junctions Are Relations of Fit

The technology layer consists of the three affordances, its unit being what the artefact can do. Following Gibson (1979), an affordance is a relation between a property of the environment and a capacity of the agent, not a feature of the tool alone.

The pedagogy layer consists of the three POA stages, and its evidence is the theory's own specification: motivating requires a gap the learner can feel, enabling requires input fitted to that gap, and assessing requires evidence a student can act on (Wen, 2015). The ability layer consists of cognitive and metacognitive language ability — decoding, vocabulary depth and parsing on one side, strategy use and comprehension monitoring on the other — instantiated for reading in Section 6.

The layers are related by fit, not by causation in series: fit between technology and stage determines the strength of the effect, fit between stage and ability its direction.

5.2 Four Propositions

Four propositions follow, each stated so that evidence could count against it.

P1, matching. The instructional value of a technology in a given stage depends on the fit between affordance and stage demand, not on model capability. *Operational form.* Capability is fixed at two levels, a general-purpose model and a smaller locally hosted one, with writing quality on a common rubric as the manipulation check; placement is the stage in which the tool is used for at least half of the stage's time, coded from lesson records. The prediction is an interaction: capability should raise scores only where the affordance matches the stage demand.

P2, mediation. The effect of AI on ability runs mainly through the three POA mechanisms, with a small direct path: output-driven motivation, input enabling and collaborative assessment, each with an established instrument. *Operational form.* A structural model with the three mechanisms as parallel mediators, in which indirect effects account for well over half of the total effect.

P3, boundary. Effects are moderated by learner self-regulation and by the quality of teacher calibration. Self-regulation is the capacity to plan, monitor and adjust comprehension without prompting, measured

with a self-regulation scale and not with a self-efficacy scale. Calibration is observed, being the proportion of machine suggestions the teacher overrides with a stated reason. *Operational form.* Multi-group or latent interaction analysis, predicting a stronger path from enabling to ability where self-regulation is high and calibration frequent.

P4, evolution. AI-POA is a variant of POA rather than a replacement, and the test is that the principles and hypotheses are unchanged. *Operational form.* Three indicators separate variant from replacement: the same ordering of stages, the same output-driven logic, with the production task set before input is supplied, and the same principles. Tempo may differ and is measured as cycles per semester and the share of stage time given to enabling. A design in which the model supplies input before the task is set counts as a different method.

P2 and P3 are within reach of current methods; P1 needs a factorial design and P4 longitudinal comparison of units.

5.3 Three Contradictions

A framework earns its keep by explaining findings that currently sit awkwardly together. Three contradictions appear, each documented in at least three studies.

The first is between studies reporting gains and studies reporting none. Gains appear in quasi-experiments that activate output or feedback (Lin et al., 2025; Ironsi & Bostanci, 2026; Zhang et al., 2024), and null or mixed results where the model produces material or items and the cycle is left unconnected (Lin & Chen, 2024; Khamrassamee et al., 2025). The positive studies activated a mediating mechanism and the others did not: whether the technology was present is not the deciding variable, which stage it entered is.

The second is between short-term reductions in anxiety and longer-term declines in motivation. The first is common in studies of a few weeks, where the model provides a low-risk interlocutor (El Hassan & Al Salwah, 2025; Nguyen, 2026); the second appears over a semester or more, where students report dependence and a loss of confidence in unaided performance (Nguyen, 2026; Li, 2024). Both can be true, and time scale is the hidden moderator.

The third is between teachers who welcome the technology and teachers who resist it. The welcome attaches to generation, which reduces preparation load (Wen, 2025a; Wu, 2026); the resistance attaches to assessment, where professional judgement appears to be at stake (Hoang & Le, 2025; El Hassan & Al Salwah, 2025). The object of the attitude is therefore a stage rather than a tool: resistance marks where the division of labour has been drawn badly.

6. Reading as the Test Case

6.1 What the Reading Evidence Shows

Ten studies take reading as their object (Table 6).

Table 6. Studies of AI in Second Language Reading

Study	Design	AI function	Outcome	Result
Wang & Yan (2022)	Design	Text adaptation and support	Comprehension	Positive, no comparison group
Wang et al. (2024)	System development	Personalised scaffolding	Use and comprehension	Positive, small sample
Lin et al. (2025)	Quasi-experiment	Pre-reading scaffolding	Comprehension, motivation, attitude	Positive
Ironsi & Bostanci (2026)	Quasi-experiment	Material generation	Comprehension, autonomy	Positive
El Hassan & Al Salwah (2025)	Survey	Open dialogue	Engagement, teacher views	Mixed
Lin & Chen (2024)	Capability test	Item generation	Item quality	Null for distractors
Khamrassamee et al. (2025)	Quasi-experiment	Diagnostic test generation	Test quality (Rasch)	Mixed
Minamitsu (2025)	Report	Translation plus model	Comprehension support	Positive, self-report
Nguyen (2026)	Survey	Open dialogue	Perceived benefit and concern	Positive with reservations
Hoang & Le (2025)	Survey	Test design	Teacher views	No outcome measured

Note. Nine of the ten are contextual studies; only Wang and Yan (2022) also combines POA with AI.

Three features stand out. The evidence is recent: nine of the ten date from 2024 or later, none predates 2022, and the designs split between three quasi-experiments and three surveys. The outcomes are not commensurable: two measure comprehension with a test, two perceived benefit, three the quality of generated items, one system behaviour, so a meta-analysis is not possible. The one study combining POA, AI and reading is a design study without a comparison group (Wang & Yan, 2022): it describes a cycle in which texts are matched to the class and support is supplied at the point of difficulty, close to what Section 5 predicts for the enabling stage, and supplies no evidence that the arrangement works.

6.2 Why reading is the hard case

Reading is where the construct is most needed and least supported, for three reasons. Comprehension is not observable: a writing gain can be inspected in a text, whereas a reading gain has to be inferred from

responses to items the model may itself have generated, and three of the ten studies use model-generated items as the outcome. Substitution is easy and invisible: a student who reads with a translation tool at hand may comprehend the passage without reading it, and the score looks the same, yet only one study (Ironsi & Bostanci, 2026) tracks tool use alongside the outcome. And the affordance most relevant to reading is the least used: reading is a process of resolving breakdowns, that affordance is dialogue, and Table 3 places dialogue in the enabling stage eight times against sixteen for generation.

6.3 What the framework predicts for reading

Applied to reading, the framework makes three predictions that differ from the ones current studies test. A metacognitive effect should precede a comprehension effect: where the technology enters the enabling stage as dialogue, the first measurable change is in monitoring and repair, so a study measuring only comprehension will read a null result for several weeks and then a gain. The dialogic cells should carry the effect: a design occupying only generative by enabling should transfer less than one occupying both, a comparison no study makes. And calibration should matter more in reading than in writing, because a machine score on a text can be checked against the text, whereas an interpretation of a misread sentence cannot. Testing these predictions needs at least one semester, a delayed post-test, a comprehension measure independent of the model, and log data on tool use during reading.

7. A Research Agenda

Five priorities follow from the gaps in Section 3.5 and the propositions in Section 5.2, ordered by feasibility rather than importance.

Mechanism, first. P2 can be tested with a structural model on a few hundred students, and the technical barrier is low.

Boundaries, second. P3 requires the same data plus a self-regulation scale and an observation protocol for calibration; multi-group and latent profile analysis are the appropriate tools.

Measurement, throughout. No instrument exists for the intensity of AI-POA implementation; two or three items per cell in Table 1 would give a fidelity scale with nine subscales, to be validated on a sample other than the one testing the hypothesis.

Matching and duration, third. P1 needs a factorial design crossing capability with placement. Interventions of one or two units cannot support claims about learning, and the absence of a design of at least one semester with a delayed post-test is the largest weakness of the evidence.

Negative effects and reading, jointly. Offloading, shallow processing and dependence should be studied as outcomes rather than acknowledged as risks, and reading is the easiest setting in which to do so, because tool use can be logged: whether the model supported reading or replaced it.

8. Limitations

The search relied on OpenAlex in place of Web of Science and Scopus. The two overlap substantially and are not equivalents: OpenAlex indexes more broadly but with less curation, its coverage of

education journals is uneven, and some relevant work is likely missing. The Chinese search was targeted rather than systematic, and only four Chinese-language items could be verified in full; work outside CSSCI and the Peking University core, and the grey literature, are absent. Five records surfaced by the acronym supplement (Appendix A) remain uncoded for want of full text. Only English and Chinese sources were included, and four fifths of the corpus dates from 2024 or later.

Coding was carried out by one person. The agreement statistic of .86 is intra-coder and measures consistency over time, not correctness; a second coder might have distributed the studies differently across stages, which is where the disagreements fell. Printing the coding frame and the study-level codes (Appendices B and C) is a partial remedy, not a substitute for double coding. Two decisions carry discretion: whether a study is core or contextual depends on whether POA shapes the design or is mentioned beside it, and assigning a study to a single cluster suppresses the fact that several use the technology in two ways.

The definition and framework are normative proposals whose value will be settled by use. The strictest constituent may be judged too demanding, since one study meets it literally; that strictness is deliberate.

9. Conclusion

The answers to the three questions are uneven in strength.

On content (RQ1), the literature describes four uses of generative AI inside a POA cycle — input, interlocutor, evaluator, design partner — unequally developed: generation and evaluation dominate, dialogue is thin, and dialogue at the motivating stage is nearly absent. The accounts converge only on the claim that the technology is useful where teacher capacity is scarce.

On development (RQ2), three periods are visible: consolidation of POA without reference to AI to 2020, first hybrids from 2021, and from 2024 a generative turn that moved the question from whether the technology fits to which stage it should enter. The concept has not developed with the literature: the vocabulary of the third period is that of the first with a tool name inserted.

On framework (RQ3), a construct can be specified, and the specification is the contribution of the paper. AI-POA is a variant of POA in which each stage is supported by at least one AI affordance, all three are represented across the cycle, and the teacher sets goals, vets material and arbitrates output. The grid decomposes it into nine observable cells, and the framework relates affordance to stage to ability through fit rather than causation.

Three claims follow: AI-POA is a variant rather than a new method, because principles and hypotheses are unchanged while tempo is not; its unit of analysis should be the stage rather than the technology, because tools turn over faster than structures; and its effectiveness depends on fit rather than capability. The reading case shows what remains — ten studies measure reading, nine without POA, one combines the two without a comparison group — so the construct is specified for a domain in which it has not been tested.

Statements

Data availability. The search strings, the coding frame and the study-level codes for all 46 studies are provided in Appendices A to C. The 1,338 screened records cannot be redistributed, but the query that generated them is given in full and can be re-executed.

Funding. This research received no external funding.

Conflicts of interest. The author declares no conflict of interest.

Author contributions. The author designed the review, conducted the search and screening, carried out the coding and appraisal, and wrote the manuscript.

Acknowledgements. The author thanks the reviewers of an earlier version for comments on the scope of the corpus and on the transparency of the coding.

Appendix A. Search Strings, Sources and Yields

Table A1. Search Strings and Records Retrieved

No	Source	String	Field	Records
S1	OpenAlex	“production-oriented approach” OR “POA”	title and abstract	602
S2	OpenAlex	(“generative AI” OR ChatGPT OR “large language model”) AND (“English reading” OR “EFL reading” OR “reading comprehension”)	title and abstract	732
S3	OpenAlex	“production-oriented approach” AND (“artificial intelligence” OR ChatGPT OR “generative AI” OR “large language model”)	title and abstract	39
C1	CNKI	(“产出导向法” OR “POA”) AND (“人工智能” OR “生成式人工智能” OR “ChatGPT” OR “大语言模型”)	topic	4 retained

Note. OpenAlex filters: title_and_abstract.search, from_publication_date 2015-01-01, to_publication_date 2026-10-31, no document-type restriction (type was inspected at screening). CNKI filters: CSSCI and Peking University core journals; the interface does not export records for this string, so the search was screened in the browser and only the retained items are counted. Search executed 19 November 2026. OpenAlex total 1,373; with the four Chinese items, 1,377 before and 1,338 after de-duplication by DOI and normalised title.

Sensitivity check on S1. The acronym form “POA” AND (“English teaching” OR “English learning” OR “language teaching”) returned 237 records, of which 185 were already in S1 and 52 were not. Title screening of the 52 left five that describe an AI component without an accessible full text: an AI-based evaluation model for college English under an output-oriented design (2024), an ESP course under AI guidance (2021), a smart teaching evaluation system (2024), a college foreign language approach built

on intelligent information technology (2021), and a unit design for an interpretation course using a virtual reality tool (2024). All five would enter the input or evaluator cluster as design proposals; they are recorded here rather than added, under the exclusion rule for records without sufficient information to code. The period begins in 2015, the year the POA system was published in its consolidated form (Wen, 2015).

Appendix B. Coding Frame

Table B1. Dimensions, categories and decision rules

Dimension	Categories	Decision rule
Stratum	Theory source, core, contextual, review	Theory source: states or revises POA and contains no AI. Core: combines POA with an AI tool in a design, implementation or evaluation. Contextual: examines AI in language learning with no POA component. Review: synthesises either literature. Assigned before cluster coding.
Design	T, D, C, Q, S, R	T theoretical or programmatic. D design, development or model construction, including a described implementation without systematic data. C case study with sustained qualitative data. Q quasi-experiment with a comparison condition. S survey or perception study. R review. Where a study reports two components, the one that carries the claim is coded.
Cluster	Input, interlocutor, evaluator, design partner, n/a	Input: the tool produces or adapts material for learners. Interlocutor: the tool exchanges turns with a learner. Evaluator: the tool scores, diagnoses or generates test items. Design partner: the tool supports the teacher's preparation. Coded from the primary function in the reported design, not from every function mentioned.
Stage	M, E, A	The POA stage in which the tool is used for a substantive part of the stage. Multiple codes allowed. A tool used only before the cycle begins is coded M.
Ability	Reading, writing, speaking, general, n/a	The language ability the study addresses. General: college English or foreign language teaching without a named ability. n/a: theory and review items with no instructional object.
Effect	Positive, mixed, null, not measured	Positive: a gain reported for the treated condition. Mixed: gains on some measures and not others, or a gain with a stated reservation. Null: no difference, or a capability test that failed on the focal criterion. Not measured: design and theory studies without outcome data.

Dimension	Categories	Decision rule
Quality	High, medium, low	Empirical: MMAT criteria for purpose, sampling, analysis and transferability. Conceptual: JBI text-and-opinion criteria for authority, analytical soundness and situatedness. High meets all applicable criteria; medium meets all but one; low meets two or fewer and is excluded.

Appendix C. Study-Level Codes

Table C1. Study-level codes for the 46 included studies

No	Study	Stratum	Cluster	Stage	Design	Ability	Quality
1	Wen (2018)	Theory	n/a	MEA	T	n/a	Medium
2	Wen (2017)	Theory	n/a	MEA	T	n/a	Medium
3	Ellis (2017)	Theory	n/a	MEA	T	n/a	Medium
4	Pu (2017)	Theory	n/a	MEA	T	n/a	Medium
5	Zhang (2020)	Theory	n/a	M	T	n/a	Medium
6	Sun (2020)	Theory	n/a	MEA	T	n/a	Medium
7	Qiu (2020)	Theory	n/a	E	T	n/a	Medium
8	Wang & Yan (2022)	Core	Input	ME	D	Reading	Medium
9	Kashyap et al. (2025)	Core	Evaluator	EA	D	Writing	Medium
10	Dong (2026)	Core	Interlocutor	E	C	Speaking	High
11	Zhao (2025)	Core	Input	E	T	Writing	Medium
12	Chen & Shen (2026)	Core	Input	ME	D	Writing	Medium
13	Zhou (2025)	Core	Input	EA	D	Writing	High
14	Zhang & Miao (2025)	Core	Input	MEA	T	General	Medium
15	Zhu (2026)	Review	n/a	n/a	R	n/a	Medium
16	Wang (2026)	Review	n/a	n/a	R	n/a	Medium
17	Zou (2026)	Core	Input	MEA	Q	Writing	High
18	Wu (2026)	Core	Design partner	ME	C	General	Medium
19	Luo & Liang (2022)	Core	Evaluator	A	Q	Writing	Medium
20	Liu & Liang (2026)	Core	Interlocutor	EA	Q	Speaking	High
21	Li (2026)	Core	Input	E	T	General	Medium
22	Wang et al. (2026)	Core	Interlocutor	EA	S	Speaking	High
23	Li (2024)	Core	Interlocutor	MEA	D	Writing	Medium
24	Zhang & Lin (2025)	Core	Evaluator	MEA	D	General	Medium

No	Study	Stratum	Cluster	Stage	Design	Ability	Quality
25	Guo (2024)	Core	Design partner	E	T	General	Medium
26	Zhang & Yang (2026)	Core	Input	MEA	D	Writing	High
27	Liu (2025)	Core	Evaluator	A	D	General	Medium
28	Zhao (2026)	Core	Design partner	E	T	Writing	Medium
29	Ma (2026)	Core	Design partner	A	T	General	Medium
30	Yuan (2026)	Core	Evaluator	MEA	D	Writing	Medium
31	Zhang et al. (2024)	Contextual	Interlocutor	EA	Q	Writing	High
32	Shi et al. (2025)	Core	Input	EA	D	Writing	Medium
33	Wen (2024)	Theory	n/a	MEA	T	n/a	Medium
34	Wang et al. (2024)	Contextual	Evaluator	E	D	Reading	Medium
35	Lin et al. (2025)	Contextual	Input	M	Q	Reading	High
36	Ironsi & Bostanci (2026)	Contextual	Input	E	Q	Reading	High
37	El Hassan & Al Salwah (2025)	Contextual	Interlocutor	E	S	Reading	High
38	Lin & Chen (2024)	Contextual	Evaluator	A	D	Reading	Medium
39	Khamrassamee et al. (2025)	Contextual	Evaluator	A	Q	Reading	High
40	Minamitsu (2025)	Contextual	Evaluator	E	D	Reading	Medium
41	Nguyen (2026)	Contextual	Interlocutor	E	S	Reading	High
42	Hoang & Le (2025)	Contextual	Evaluator	A	S	Reading	Medium
43	Wen (2025a)	Core	Design partner	ME	T	General	Medium
44	Zhang (2025)	Core	Design partner	ME	C	General	Medium
45	Qiu (2025)	Core	Interlocutor	E	T	General	Medium
46	Wen (2025b)	Core	Design partner	MEA	T	General	Medium

Note. Stage: M motivating, E enabling, A assessing; a study may occupy more than one. Design: T theoretical, D design or development, C case study, Q quasi-experiment, S survey, R review. Ability: the language ability the study addresses, not the outcome it measures. Quality: MMAT for empirical studies and JBI text-and-opinion for conceptual ones; the six studies rated low were excluded before coding and do not appear.

Appendix D. PRISMA 2020 Checklist

Table D1. PRISMA 2020 checklist

No	Checklist item	Reported at
1	Title identifies the report as a systematic review	Title
2	Structured abstract	Abstract
3	Rationale	1.1, 1.2
4	Objectives	1.3
5	Eligibility criteria	2.3, Appendix B
6	Information sources	2.2, Appendix A
7	Full search strategy	2.2, Table A1
8	Selection process	2.4, Figure 1
9	Data collection process	2.5
10	Data items	2.5, Appendix B
11	Risk of bias assessment	2.5
12	Effect measures	2.5 (direction only; no pooled estimate)
13	Synthesis methods	2.6
14	Reporting bias assessment	3.5, 8
15	Certainty assessment	3.4, 8
16	Study selection	2.4, Figure 1
17	Study characteristics	Table 2, Appendix C
18	Risk of bias in studies	2.5, Appendix C
19	Results of individual studies	3.2 to 3.4, Table 6
20	Results of syntheses	3.2 to 3.4, Tables 3 to 5
21	Reporting biases	3.5
22	Certainty of evidence	3.4, 8
23	Discussion	4 to 7
24	Registration and protocol	Not registered; no protocol published
25	Support	Statements
26	Competing interests	Statements
27	Availability of data and code	Statements, Appendices A to C

Note. The review was not registered and no protocol was published before the search. For a review that makes claims about the state of a field, that is a weakness rather than a formality: a registered protocol would have fixed the inclusion criteria before the corpus was known, and the criteria in Section 2.3 were refined during screening.

References

Note. An asterisk marks the 46 studies included in the review.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Chen, H. F., & Shen, L. J. (2026). Generative artificial intelligence and teaching according to aptitude in private undergraduate college English: A production-oriented approach framework. *International Journal of Academic Research in Business and Social Sciences*, *16*(10).

Dong, H. (2026). Case study on the impact of production-oriented approach empowered by artificial intelligence on IELTS speaking performance. *Pacific International Journal*, *9*(3). https://doi.org/10.55014/pij.v9i3.1023

El Hassan, F. A. M., & Al Salwah, A. F. (2025). Exploring the impact of ChatGPT on EFL reading practices: Opportunities and challenges. *International Journal of English Language Teaching*, *13*(1), 85-93.

Ellis, R. (2017). The production-oriented approach: Moving forward. *Chinese Journal of Applied Linguistics*, *40*(4), 454-458.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Guo, J. J. (2024). Practical paths for college English teaching in the context of digital transformation. *Advances in Social Science, Education and Humanities Research*. https://doi.org/10.2991/978-2-38476-265-1_43

Hoang, H. P., & Le, T. H. Q. (2025). High school teachers' perceptions of using ChatGPT for designing English reading comprehension tests. *Tay Nguyen Journal of Science*. https://doi.org/10.5281/zenodo.17538660

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Carvalho, A., et al. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285-291.

Ironsi, C. S., & Bostanci, H. B. (2026). Enhancing the reading comprehension and autonomy of foreign language students using generative AI. *Frontiers in Education*. https://doi.org/10.3389/feduc.2026.1763238

Kashyap, A., Nag, S., & Sharma, A. (2025). College English teaching model following production-oriented approach under artificial intelligence. In *Proceedings of the International Conference for Emerging Technology*. https://doi.org/10.1109/INCET64471.2025.11140893

Khamrassamee, P., Junpeng, P., & Intharah, T. (2025). The effectiveness of using large language model to generate English reading comprehension diagnostic tests. In *Proceedings of the International Technical Conference on Circuits/Systems, Computers and Communications*. https://doi.org/10.1109/ITC-CSCC66376.2025.11137621

Li, H. (2026). Constructing a blended foreign language teaching model empowered by generative artificial intelligence. *English Language Teaching and Linguistics Studies*, *8*(4). https://doi.org/10.22158/eltls.v8n4p131

Li, L. (2024). Empowering college English writing instruction with ChatGPT: Integration of the teacher-AI-student framework. *Trends in Social Sciences and Humanities*. https://doi.org/10.61784/tsshr3125

Lin, C. C., Lin, T. H., & Tang, C. K. (2025). Enhancing English reading comprehension, learning motivation and attitude through AI-supported pre-reading scaffolding. *Journal of Computer Assisted Learning*, *41*(6), e70150.

Lin, Z. Q., & Chen, H. L. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, *123*, 103344.

Liu, Y. (2025). The innovative construction of a multidimensional dynamic evaluation system for college English courses. *Advances in Social Science, Education and Humanities Research*. https://doi.org/10.2991/978-2-38476-487-7_17

Liu, Y., & Liang, B. (2026). Higher-order thinking and English public speaking in an AI-empowered production-oriented approach. *SAGE Open*. https://doi.org/10.1177/21582440261501787

Luo, Y., & Liang, W. (2022). Application of POA in teaching mode of college English writing based on artificial intelligence. In *Proceedings of the 2022 IEEE 5th Eurasian Conference on Educational Innovation (ECEI)* (pp. 105-109). IEEE. https://doi.org/10.1109/ECEI53102.2022.9829466

Ma, D. (2026). An OBE-POA framework for AI-enhanced collaborative development: A multidimensional competency model. In *Smart Innovation, Systems and Technologies*. https://doi.org/10.1007/978-3-032-23712-5_18

McArthur, A., Klugarova, J., Yan, H., & Florescu, S. (2020). Systematic review of text and opinion. In *JBI Manual for Evidence Synthesis*. JBI.

Minamitsu, Y. (2025). *English reading comprehension support through machine translation and generative AI* [Research report].

Nguyen, T. X. C. (2026). Students' perceptions towards the use of ChatGPT in English reading comprehension. *International Journal of Education, Social Science and Research*. https://doi.org/10.37500/ijessr.2026.9522

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.

Pu, S. (2017). Appraisal of the production-oriented approach: An introduction. *Chinese Journal of Applied Linguistics*, *40*(4), 365-372.

Qiu, L. (2020). Enabling in the production-oriented approach: Theoretical principles and classroom implementation. *Chinese Journal of Applied Linguistics*, *43*(3), 285-304.

Qiu, L. (2025). AI-empowered enabling instruction in the production-oriented approach. *Modern Foreign Languages*, *48*(5), 690-700. [in Chinese]

Shi, X., Wu, W., & Zhang, Y. (2025). Teaching reform practice of professional English in intelligent manufacturing under the "AI + POA" model. In *Proceedings of the International Conference on Artificial Intelligence*. https://doi.org/10.1145/3756423.3756556

Sun, S. G. (2020). The production-oriented approach updated: Introduction to the special issue. *Chinese Journal of Applied Linguistics*, *43*(3), 259-267.

Sun, S., & Qi, T. (2025). Teacher-AI-student collaborative assessment in the interpreting classroom. *Modern Foreign Languages*, *48*(5), 701-712. [in Chinese]

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*, 45.

Wang, J., Ismail, H. H., & Mansor, A. Z. (2026). Assessing Chinese EFL learners' engagement and speaking proficiency in ChatGPT-integrated POA. In *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-3-032-36012-0_13

Wang, X. L. (2026). The theoretical evolution, knowledge structure, and digital-intelligent turn of the production-oriented approach. *Lecture Notes in Education Psychology and Public Media*. https://doi.org/10.54254/2753-7048/2026.38275

Wang, X. Z., Zhong, Y. H., & Huang, C. Q. (2024). ChatPRCS: A personalized support system for English reading comprehension based on ChatGPT. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2024.3405747

Wang, X., & Yan, K. (2022). College English reading teaching integrating production oriented approach from the perspective of artificial intelligence. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/3411203

Wen, Q. (2015). Constructing the theoretical system of the production-oriented approach. *Foreign Language Teaching and Research*, *47*(4), 547-558. [in Chinese]

Wen, Q. (2016). Teacher-student collaborative assessment: A new form of assessment created by the production-oriented approach. *Foreign Language World*, (5), 37-43. [in Chinese]

Wen, Q. (2017). The production-oriented approach. In *The Routledge handbook of English language teaching*. Routledge.

Wen, Q. (2018). The production-oriented approach to teaching university students English in China. *Language Teaching*, *51*(4), 526-540.

Wen, Q. (2020). *The production-oriented approach: Exploring a foreign language education theory with Chinese characteristics*. Foreign Language Teaching and Research Press. [in Chinese]

Wen, Q. (2024). Reconsidering the POA. In *English language education in China*. Springer. https://doi.org/10.1007/978-3-031-65080-2_12

Wen, Q. (2025a). Exploratory research on AI-empowered POA English textbook compilation. *Modern Foreign Languages*, *48*(5), 667-678. https://doi.org/10.20071/j.cnki.xdwy.20250620.007 [in Chinese]

Wen, Q. (2025b). AI-empowered reform of POA foreign language teaching: An innovative exploration of the shift to immediate communication. *Foreign Language Learning Theory and Practice*, (4), 40-49. [in Chinese]

Wu, T. (2026). AI empowered textbook use under the production-oriented approach: A case study. *Scientific Journal of Humanities and Social Sciences*. https://doi.org/10.54691/pefhgv06

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93-112.

Yuan, Y. (2026). Exploration and practice of AI-empowered smart teaching model for college English. *Frontiers in Educational Research*, *9*(4). https://doi.org/10.25236/fer.2026.090409

Zhang, H. Q., & Yang, S. R. (2026). Exploration of the POA + 5E model in English teaching for art postgraduates empowered by AI. *Journal of Literature and Art Studies*. https://doi.org/10.17265/2159-5836/2026.02.003

Zhang, L. L. (2020). Motivating in the production-oriented approach: From theory to practice. *Chinese Journal of Applied Linguistics, 43*(3), 268-284.

Zhang, M., Wang, Y., & Shang, Q. (2024). Using AI chatbots to improve university EFL students' speech scripts' writing quality and critical thinking. In *Proceedings of the International CALL Research Conference*. https://doi.org/10.29140/9780648184485-52

Zhang, W. (2025). Classroom practice of AI-empowered POA textbook use. *Modern Foreign Languages*, *48*(5), 679-689. [in Chinese]

Zhang, X. L., & Lin, Q. R. (2025). An AI-powered instructional design and evaluation framework based on POA and ChatGPT. *Frontiers in Artificial Intelligence and Applications*. https://doi.org/10.3233/FAIA250440

Zhang, Y. S., & Miao, M. (2025). The application of generative artificial intelligence in POA guided foreign language teaching. *Advances in Social Science, Education and Humanities Research*. https://doi.org/10.2991/978-2-38476-364-1_55

Zhao, R. (2025). Application of artificial intelligence technology in enhancing the effectiveness of production-oriented approach to literacy instruction and instructional design. *Journal of Combinatorial Mathematics and Combinatorial Computing*. https://doi.org/10.61091/jcmcc127a-326

Zhao, S. Y. (2026). Teaching design of AI-enabling junior high school English instruction under the POA model. *English Language Teaching and Linguistics Studies*, *8*(3). https://doi.org/10.22158/eltls.v8n3p57

Zhou, M. S. (2025). An innovative study on college English writing courses based on the POA theory. *Advances in Social Science, Education and Humanities Research*. https://doi.org/10.2991/978-2-38476-444-0_51

Zhu, Q. (2026). A thematic review and synthesis of empirical studies connecting the production-oriented approach and generative artificial intelligence. *Open Journal of Modern Linguistics*, *16*(2). https://doi.org/10.4236/ojml.2026.162010

Zou, M. (2026). Research on the application of the production-oriented approach in junior high school English teaching. *Creative Education Studies*, *14*(7). https://doi.org/10.12677/ces.2026.147522