HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

AI-BASED ADAPTIVE LEARNING IN ONLINE FOREIGN LANGUAGE EDUCATION

Majidov Farohiddin

Independent researcher of

Samarkand State Institute of Foreign Languages

farohiddinmajidov21@gmail.com

https://orcid.org/0009-0001-3980-8117

Uzbekistan, Samarkand.

Doi: https://doi.org/10.5281/zenodo.23228687

Abstract. This paper explains how artificial intelligence can make online language courses adapt to each learner, and what this means for teaching German. Only studies published between 2022 and 2026 were reviewed. Recent research shows that adaptive systems help learners: across 45 studies the average benefit was g = 0.70, and across 41 studies of AI chatbots it was 0.576. New scheduling programs that decide when a learner should review a word have also become much more accurate, cutting prediction errors by up to 64% on very large datasets. However, most of this research does not say what an adaptive program should actually track in a specific language. This paper therefore does two things. First, it shows that adaptivity is not one decision but six: choosing the next task, ordering tasks, giving feedback, planning review, placing learners at a level, and defining what can be practised. Second, it turns eight common German problems — article gender, plurals, cases after two-way prepositions, separable verbs, adjective endings, the auxiliary in the Perfekt, word order in subordinate clauses, and long compound nouns — into clear rules that a computer program can follow, each with a signal that starts extra practice and a condition that ends it. A twelve-week study with 120 learners at A2–B1 level is proposed to test these rules. The main conclusion is simple: adaptive technology only helps if it knows what is difficult in the language being taught.

Keywords: adaptive learning, artificial intelligence, online language education, German as a foreign language, learner model, intelligent tutoring, CEFR, research design.

Introduction. Online foreign language education has moved from delivering the same sequence of materials to every learner towards systems that change their behaviour in response to what an individual learner does. The shift is technological in its mechanism but didactic in its purpose: an adaptive system attempts to keep each learner in the narrow band where tasks are neither trivial nor impossible, and to spend instructional time on exactly those forms the learner has not yet stabilised. Between 2022 and 2026 the empirical case for this approach became considerably firmer. Wang and colleagues synthesised 45 independent studies and found an overall effect of g = 0.70 for AI-enabled adaptive learning systems on cognitive learning outcomes when compared with non-adaptive interventions; notably, the type of AI used in the adaptive engine did not moderate the effect, whereas duration, discipline and the adaptive target did [1]. For conversational applications, Lyu, Lai and Guo pooled 31 studies with 41 effect sizes and 2,943 participants and obtained g = 0.608 with substantial heterogeneity (I^2 = 80.75%) [2], while Li, Wang and Yang synthesised 41 generative-AI studies with

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

48 effect sizes and 3,515 participants and obtained 0.576, 95% CI [0.385, 0.768], with vocabulary emerging as one of the strongest moderators [3].

The heterogeneity is the interesting part. It indicates that the label “adaptive” covers systems whose internal decisions differ profoundly, and that effects depend on what is adapted rather than on whether adaptation is present. This is where the pedagogical literature on online language teaching remains underspecified. Recent reviews of AI in second-language learning describe outcomes and moderators [4], and studies of digital assistants in German as a foreign language describe opportunities and risks [5], but few publications state, in a form a developer or a course designer could implement, which linguistic decisions the engine has to make in a given target language.

German is a productive language for making this concrete. Its difficulty for learners is concentrated in a small number of recurrent, formally describable systems, and each of them has been studied within the last four years. Grammatical gender remains problematic even under focused instruction: Henry taught 68 first-semester learners 24 target nouns balanced across the three genders and compared Processing Instruction with traditional instruction and with categorisation-memorisation, retesting after a four-week delay [6]. Gender assignment interacts with compounding, since determiner priming with der, die and das shows that the gender of a German noun-noun compound is driven by its morphological head, a result obtained across two experiments with 180 native speakers [8]. Particle-verb separation is not freely productive but lexically restricted and formulaic [7], and compound processing in learners depends on the form of the first-language translation, as shown for 52 German compounds tested with Czech-speaking learners [9]. Learner corpora of German continue to grow: a longitudinal study of 30 students over four years annotated 120 texts within a learner corpus of more than 5,700 texts and documented non-linear approximation to target-language patterns [20].

Literature review and research methodology. The literature was assembled through a structured search of peer-reviewed journals, conference proceedings, technical reports of system developers and examination-provider documentation, restricted to items published or last updated between 2022 and 2026. Twenty-one sources were retained under three inclusion criteria: the source reports either pooled quantitative evidence on adaptive or AI-supported language learning, or a documented algorithmic mechanism of adaptation, or recent data on German learner language and German assessment practice. Advocacy pieces without data or mechanism were excluded, as were pre-2022 publications, so that the synthesis reflects the current state of the field rather than its history.

Four strands emerge. The first is effect-size evidence. Beyond the three meta-analyses already cited [1]–[3], a systematic review of AI-driven intelligent tutoring systems confirms positive but context-dependent outcomes [17]; a meta-analysis of AI-assisted vocabulary teaching covering 18 studies, 23 effect sizes and 1,823 participants reports $g = 1.309$, 95% CI [0.876, 1.742], with $I^2 = 93.868\%$ and $\kappa = 0.92$ for study selection [15]; and mobile-assisted vocabulary research illustrates how strict inclusion criteria reduce a pool of 10,163 records to 65 analysable studies [16]. The second strand is algorithmic. Su and colleagues modelled the dynamics of memory on a released benchmark of 220 million review rows and reduced prediction and scheduling error by 64% while lowering cost by 17% relative to established baselines [10]. Xiao and Wang, working on 71,697 processed learner groups drawn from the same order of data, reported a mean absolute error of 0.0274 for recall prediction against 0.1070 for half-life regression and 0.4535 for a Leitner schedule, and average recall probabilities of 0.920 and 0.942 in two simulated environments [11]; a further AIED contribution applies personalised spaced-repetition scheduling directly to language learning [12]. Developer documentation describes an operational student model estimating learner proficiency and item

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

difficulty from more than one billion exercises completed daily, and using those estimates to sequence exercises, decrease the difficulty of persistently failed items and schedule review [19]. The third strand is assessment: adaptive testing is operational in German, since the Goethe-Test PRO adjusts automatically to the test taker and reports a CEFR level from A1 to C2 [18], [21]; item parameters can be predicted from linguistic features before live piloting [14]; and large language models have been used to make adaptive item selection explainable, evaluated with AUC, accuracy and exposure rate [13]. The fourth strand is German learner language: gender acquisition under instruction [6], head-driven gender in compounds [8], particle-verb separation [7], learner compound processing [9] and recent error-annotated German learner corpora [20].

The methodology has three components. (i) Structured qualitative synthesis: each reviewed system or study was coded on adaptive source (what learner evidence is used), adaptive target (what the system changes), granularity (item, task, unit or course) and evidence type. (ii) Design-analytic mapping: for each German difficulty cluster, three specifications were derived — the diagnostic signal that identifies the gap, the adaptive action to be triggered, and the exit condition that returns the learner to the main path. Clusters were selected because they are attested in German learner research published since 2022 and are formally detectable in short written or selected-response productions. (iii) Protocol specification: a quasi-experimental design was drafted in a form that allows replication in an ordinary online course, with parameters chosen to satisfy the inclusion thresholds used in current meta-analytic practice, namely a control group, a minimum of 15 participants per group and a treatment of at least ten weeks [16].

Two methodological limits follow from this design. Coding of adaptive source and target was performed by a single analyst, so no interrater reliability can be reported, whereas comparable syntheses report agreement of 90% or higher and formal kappa coefficients [15], [16]. And the rule table is a design artefact: it is internally consistent with the cited German learner data, but its instructional value remains a hypothesis until the protocol in Section III is executed.

Results. Table I reports the synthesis of adaptive mechanisms across the reviewed sources.

Table I. Synthesis of adaptive layers in the reviewed systems (2022–2026)

Adaptive layer	Learner evidence used	System action	Evidence
Item selection	Estimated ability and item difficulty	Choose next item near ability level	[13], [14], [19]
Sequencing	Accuracy within the current session	Raise or lower difficulty; resurface item	[19]
Feedback	Error type in the learner response	Hint, then targeted explanation	[5], [19]
Review scheduling	Recall history and item features	Predict recall probability; set interval	[10], [11], [12]
Placement	Adaptive test performance	Assign CEFR level A1–C2	[18], [21]

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Domain modelling	Error-annotated learner corpus	Define the inventory of trainable forms	[7], [20]

The synthesis shows that adaptivity is not a single decision but a stack. Review scheduling is the layer with the strongest algorithmic evidence, since the 2023–2025 models were trained and validated on benchmarks of hundreds of millions of recall events and report explicit error reductions [10], [11]; feedback is the layer with the weakest formal specification, because error interpretation requires language-specific knowledge that generic difficulty estimates do not contain. This is precisely the gap that the German rule table in Table II addresses.

Table II. German difficulty clusters as adaptive decision rules.

German difficulty cluster	**Diagnostic signal**	**Adaptive action**	**Exit condition**
Grammatical gender (der/die/das)	Article errors clustered in one gender class	Insert items grouped by gender class, with delayed retesting [6]	85% article accuracy over 20 items
Plural formation	Plural forms overgeneralised to one pattern	Contrast competing patterns (Tisch / Tische, Kind / Kinder, Frau / Frauen, Auto / Autos)	Correct plural in 8 of 10 new nouns
Case after two-way prepositions	Accusative and dative confused after in, auf, über	Motion-versus-location minimal pairs (in die Schule / in der Schule)	Both members correct in 5 pairs
Particle-verb separation	Particle left attached in main clauses	Frame drills with high-frequency particle verbs (aufstehen: ich stehe um sieben auf) [7]	Correct separation in 6 of 8 items
Adjective endings	Endings correct after der but not after ein	Determiner-conditioned tables (der gute Wein / ein guter Wein)	80% accuracy across three determiner types
Auxiliary in the Perfekt	haben generalised to motion verbs	Verb-class sorting (ich habe gearbeitet / ich bin gefahren)	Correct auxiliary in 9 of 10 items
Subordinate word order	Finite verb not in final position after dass, weil	Sentence-reordering tasks (…, weil ich Deutsch lerne)	Verb-final in 5 consecutive productions

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

German difficulty cluster	Diagnostic signal	Adaptive action	Exit condition
Nominal compounding	Compounds decoded as unknown single units	Head-based segmentation training (Krankenversicherungskarte) [8], [9]	Correct segmentation of 4 novel compounds

Table II makes the central design claim operational. Each row converts a documented learner problem into three machine-readable elements. The gender row reflects the finding that instructional technique measurably changes gender outcomes and that gains must be verified after a delay rather than immediately [6]. The compounding row reflects determiner-priming evidence that compound gender is assigned by the morphological head [8], which means segmentation training should isolate the head first, and learner evidence that compound processing is mediated by the first-language translation form [9]. The particle-verb row rests on corpus evidence that separation occurs in lexically sparse and formulaic contexts while joint use is lexically diverse [7]; an adaptive engine should therefore practise separation in fixed frames rather than in free production. The exit conditions are deliberately expressed as observable thresholds so that the engine can leave a remedial branch without teacher intervention.

Table III. Proposed evaluation protocol

Protocol element	Specification
Design	Quasi-experimental, two groups, pre-test / post-test / delayed post-test after four weeks [6]
Participants	120 online learners of German at A2 entering B1; 60 per group; minimum 15 per group per current meta-analytic practice [16]
Duration	12 weeks, 3 sessions of 25 minutes per week
Treatment	Adaptive condition: item selection, sequencing, feedback and review driven by the rules in Table II
Control	Same content and time, fixed sequence, non-adaptive review
Placement	Adaptive CEFR placement before assignment [18]
Outcome measures	Cluster-specific accuracy (gender, plural, case, particle verbs); written production scored with a CEFR-referenced rubric [20], [21]
Process measures	Items attempted, time on task, predicted recall error, branch entries and exits [11]
Analysis	ANCOVA on post-test with pre-test as covariate; Hedges's g with 95% CI; mixed-effects logistic model on item-level responses
Benchmark	Effects compared against g = 0.70 [1], g = 0.608 [2] and 0.576 [3]

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Discussion. The results support a specific reading of the recent meta-analytic record. If the type of AI in the adaptive engine does not moderate outcomes while the adaptive target does [1], then the engineering choice matters less than the didactic choice of what to adapt. The algorithmic literature of the same period points the same way: the reported error reduction of 64% [10] and the drop in mean absolute error from 0.4535 under a Leitner schedule to 0.0274 under a memory-dynamics model [11] were achieved by modelling recall of specific items with item-level features, not by a more powerful learner classifier in the abstract.

This has a direct consequence for online German courses. A generic engine that records a single scalar "ability" cannot distinguish a learner who fails an item because the noun's gender is unknown from one who fails because the two-way preposition triggered the wrong case; both events lower the same number. The rule table in Table II replaces one scalar with eight diagnosable states, each with its own remedial branch. The cost is knowledge engineering; the benefit is that feedback can name the rule that was violated, which is what learners in German as a foreign language settings expect from a digital assistant and what current generative tools deliver inconsistently [5].

Three risks deserve explicit statement. First, over-adaptation. If the engine always keeps difficulty near the current ability estimate, learners may never encounter the desirable difficulty that consolidates retention; the documented practice of lowering the difficulty of persistently failed items [19] must therefore be paired with scheduled re-exposure, and gains must be verified on a delayed post-test, as in the German gender study that retested after four weeks [6]. Second, construct narrowing. Cluster-specific accuracy is easy to measure, and there is a real danger that an adaptive course optimises measurable morphology while neglecting pragmatic and discourse competence, which is why the protocol in Table III retains a rubric-scored production measure alongside item-level metrics [20], [21]. Third, data dependence. Adaptive estimates are only stable with sufficient item exposure, which is why item parameters for adaptive language tests are increasingly predicted from linguistic features before live data exist [14] and why explainability has become a research target in adaptive testing [13]; a small institutional course will not reach benchmark data volumes of hundreds of millions of events [10], [11] and should borrow parameters rather than estimate them from scratch.

Conclusion. Adaptive learning in online foreign language education is best understood as a stack of four decisions — item selection, sequencing, feedback and review scheduling — supplemented by adaptive placement and by a domain model derived from learner data. Evidence published between 2022 and 2026 is favourable ($g = 0.70$ across 45 adaptive-system studies; $g = 0.608$ across 31 chatbot studies; 0.576 across 41 generative-AI studies) but heterogeneous, and the engine's algorithm matters less than what it is asked to adapt. For German, adaptivity can be specified precisely: eight recurrent difficulty clusters were converted into diagnostic signals, adaptive actions and observable exit conditions, each anchored in German learner research from the last four years, and a twelve-week quasi-experimental protocol with 120 A2–B1 learners was drafted to test them. The contribution of the article is this bridge between a favourable but coarse effect-size literature and the language-specific decisions an online German course must implement.

REFERENCES:

1. X. Wang, R. T. Huang, M. Sommer, B. Pei, P. Shidfar, M. S. Rehman, A. D. Ritzhaupt, and F. Martin, "The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis," Journal of Educational Computing Research, 2024,

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

doi: 10.1177/07356331241240459. [Online]. Available: https://journals.sagepub.com/doi/10.1177/07356331241240459

2. B. Lyu, C. Lai, and J. Guo, "Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies," International Journal of Applied Linguistics, vol. 35, no. 2, pp. 834–851, May 2025, doi: 10.1111/ijal.12668. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/ijal.12668
3. M. Li, Y. Wang, and X. Yang, "Can generative AI chatbots promote second language acquisition? A meta-analysis," Journal of Computer Assisted Learning, vol. 41, no. 4, art. e70060, Aug. 2025. [Online]. Available: https://eric.ed.gov/?id=EJ1478367
4. X.-Y. Wu, "Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators," System, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0346251X2400280X
5. M. Doustizadeh and S. H. Golestaneh, "Artificial intelligence in teaching German as foreign language: Exploring the opportunities and challenges of digital educational assistants," Critical Language and Literary Studies, vol. 22, no. 35, pp. 146–165, 2025, doi: 10.48308/clls.2025.240094.1354. [Online]. Available: https://clls.sbu.ac.ir/article_105948.html
6. N. Henry, "The effects of Processing Instruction on the acquisition and processing of grammatical gender in German," Language Teaching Research, 2022, doi: 10.1177/13621688221096368. [Online]. Available: https://journals.sagepub.com/doi/10.1177/13621688221096368
7. D. Jach and J. Zeng, "Analysing the use of particle verbs in German as a foreign language: Unproductive particle separation and a proposal for a data-driven learning intervention," Language Teaching Research, 2024, doi: 10.1177/13621688241300399. [Online]. Available: https://journals.sagepub.com/doi/10.1177/13621688241300399
8. A. Lorenz and E. Beyersmann, "Head-driven gender assignment in noun-noun compound recognition: Evidence from a determiner priming task in German," PLoS One, vol. 21, no. 5, art. e0348377, 2026, doi: 10.1371/journal.pone.0348377. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC13189298/
9. A. Kříž and D. Bordag, "The role of L1 translation form in L2 compound processing: The case of native Czech speakers processing German noun-noun compounds," Frontiers in Communication, vol. 8, art. 1133927, 2023, doi: 10.3389/fcomm.2023.1133927. [Online]. Available: https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2023.1133927/full
10. J. Su, J. Ye, L. Nie, Y. Cao, and Y. Chen, "Optimizing spaced repetition schedule by capturing the dynamics of memory," IEEE Transactions on Knowledge and Data Engineering, vol. 35, pp. 10085–10097, Oct. 2023, doi: 10.1109/TKDE.2023.3251721. [Online]. Available: https://www.computer.org/csdl/journal/tk/2023/10/10059206/1LiKKhcXw52
11. Q. Xiao and J. Wang, "DRL-SRS: A deep reinforcement learning approach for optimizing spaced repetition scheduling," Applied Sciences, vol. 14, no. 13, art. 5591, 2024, doi: 10.3390/app14135591. [Online]. Available: https://ira.lib.polyu.edu.hk/bitstream/10397/111924/1/applsci-14-05591.pdf
12. B. Ma, S. Fukui, Y. Ando, and S. Konomi, "Personalized language learning using spaced repetition scheduling," in Artificial Intelligence in Education (AIED 2025), Lecture Notes in Computer Science, vol. 15880, Cham, Switzerland: Springer, 2025, doi: 10.1007/978-3-031-98459-4_19.
13. C. Cheng, G. Zhao, Z. Huang, Y. Zhuang, Z. Pan, Q. Liu, X. Li, and E. Chen, "Towards explainable computerized adaptive testing with large language model," in Findings of the

Association for Computational Linguistics: EMNLP 2024, pp. 2655–2672. [Online]. Available: https://aclanthology.org/2024.findings-emnlp.149.pdf
14. "Accelerating item piloting with BERT embeddings," in Proc. 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2024. [Online]. Available: https://aclanthology.org/2024.bea-1.35.pdf
15. B. Şimşek, B. Direkci, B. Koparan, S. Akbulut, E. E. Şimşek, M. Gülmez, and D. N. Ökmen, "Artificial intelligence-assisted vocabulary teaching: A meta-analysis study," Frontiers in Artificial Intelligence, vol. 9, art. 1806700, 2026, doi: 10.3389/frai.2026.1806700. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC13149277/
16. Y. Zhou and M. Zhou, "A meta-analysis on mobile-assisted vocabulary learning: Do mobile applications help?," ReCALL, 2025, doi: 10.1017/S0958344025100335. [Online]. Available: https://www.cambridge.org/core/journals/recall/article/79C8F00E48D521ED3BCFE0625B5549F3
17. "A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education," npj Science of Learning, 2025. [Online]. Available: https://www.nature.com/articles/s41539-025-00320-7
18. Goethe-Institut, "Goethe-Test PRO: German for professionals." [Online]. Available: https://www.goethe.de/ins/se/en/spr/prf/bulats.cfm (accessed Sep. 5, 2026).
19. C. Freeman, A. Kittredge, H. Wilson, and B. Pajak, "The Duolingo method for app-based teaching and learning," Duolingo Research Report, Jan. 2023. [Online]. Available: https://duolingo-papers.s3.amazonaws.com/reports/Duolingo_whitepaper_duolingo_method_2023.pd
20. H. Huang and Y. Li, "Zum Nachfelderwerb im Schriftdeutschen bei chinesischen Deutschlernenden: Eine korpusgestützte longitudinale Untersuchung," Korpora Deutsch als Fremdsprache (KorDaF), 2024. [Online]. Available: https://kordaf.tujournals.ulb.tu-darmstadt.de/article/id/4145/
21. Goethe-Institut, "German exams A1–C2." [Online]. Available: https://www.goethe.de/en/spr/prf.html (accessed Sep. 5, 2026).