



Listen. Answer. Memorize.

Building and Running Verified Learning Paths for Geoscience Self-Study with an AI Assistant

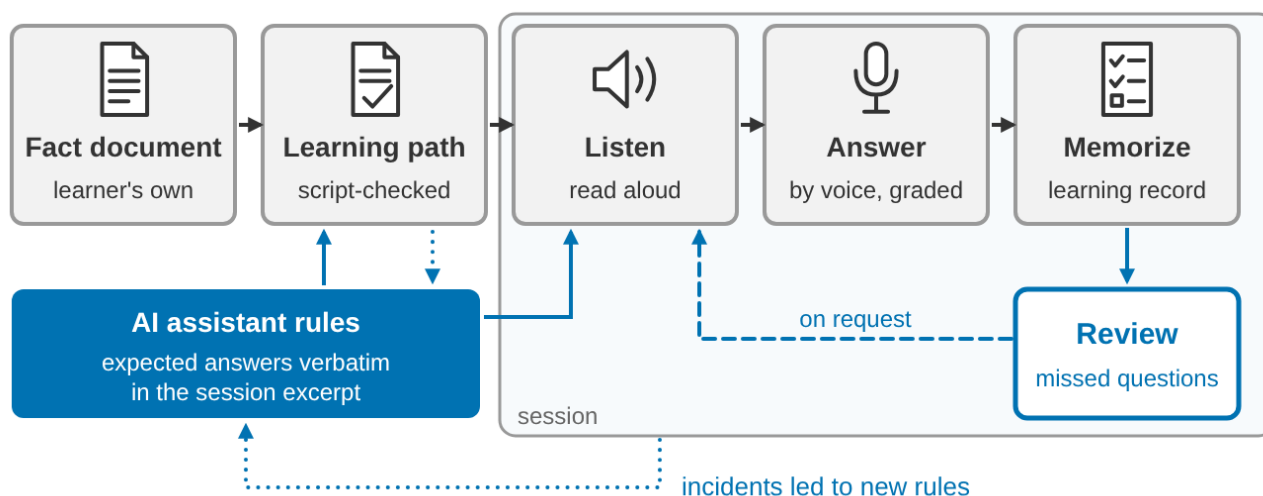
Stefan Mohr

5 Independent Researcher, Hamburg, Germany. ORCID 0000-0003-4657-3501

Abstract. AI tutors built on large language models can state things that no source supports, a risk that weighs on self-study. This paper describes how a single adult learner built, with an AI assistant, learning paths from his own verified fact documents in the geosciences and, for two of the five cases, learned them in voice calls. The skill lernpfad-sitzungen cuts a
10 fact document into sessions of whole paragraphs with three questions each, whose expected answers must appear word for word in the excerpt of the session; a check script verifies this. In a session, the assistant reads the excerpt aloud, asks the questions, grades the spoken answers tolerantly and keeps a learning record, from which review sessions are built on request. The study covers five cases: four learning projects, on southern Norway, Earth's geological eras, the geology of Germany and volcanology, and a narrative version of the fact document on the geological eras. All six learning paths checked, among
15 them two versions of the path on southern Norway, passed the check script, three of them only after chapters that are not learning material had been named as exclusions. By 13 October 2026, the learning records held 25 sessions on two cases, 23 of them entirely in a voice call, and two review sessions had been completed. By the learner's own account, sessions of five to ten minutes let him learn on the move and in short breaks. Incidents in building learning paths and in sessions produced new rules for the assistant; in the narrative version, for example, all questions were asked as worded only once they were
20 shown in the text chat before each call. The study is a self-experiment without a control condition and makes no claim about retention. The rules of the skill contain no subject content; in the author's assessment, it should also suit other subjects that rely on factual and conceptual knowledge, which has not been tested.



Graphical Abstract



25 1 Introduction

In *Talk, Check, Decide* (Mohr, 2026), the author described a workflow in which an AI assistant helps him write and check geological excursion reports, working almost entirely by voice. Building skills with which fact documents are written and checked led to a further wish: to have their content taught as well. This paper takes that step. It argues that a learner can use AI to assemble, for himself, learning content that is bound word for word to a checked fact document, and to have it taught to him in different ways; how good that content is depends on how the fact document itself was checked (Sect. 4.1). An AI tutor, however, can state things that its source does not contain or contradict what it does contain (Ji et al., 2023). The learning environment studied here therefore binds every session to a fact document that the learner supplies: the excerpt is read out word for word, and every expected answer must occur word for word in it. The study is a self-experiment in which the author is both the learner and the developer of the learning environment.

Geoscience material makes learning by ear particularly demanding, because it is dense with time spans, ages and units. Learners often misjudge the ages of geological events by orders of magnitude (Cheek, LaDue and Shipley, 2017), and many adults treat thousand, million and billion as evenly spaced steps rather than as factors of a thousand (Landy, Silbert and Goldin, 2013). In the sessions studied here, the learner confused the order of magnitude of ages (millions instead of billions of years), once gave up on a question with the words ‘I don’t know, those were too many numbers’ (translated from German), and the speech recognition turned ‘55 Ma’ into the German for ‘55 times’. From version 2.5.0 on, the skill aims its questions at relations and events instead of numbers, dates and names.

This paper addresses three research questions:



RQ1 How can a learner build, with an AI assistant, verified fact documents and learning paths from them, so that every question in a session can be answered verbatim from the fact document?

45 RQ2 How do learning sessions run when the excerpt is read aloud to the learner in a voice call and the learner answers the questions by voice, and which rules for the assistant arose from incidents in real sessions?

RQ3 How does a learning record kept by the AI assistant track what the learner has not yet mastered, and how are review sessions built from it?

The three words of the title name the three parts of a session (Sect. 3.5.1 to 3.5.3); ‘Memorize’ stands for the learning record
 50 and the review sessions, not for measured retention.

2 Related work

Retrieval practice and spacing are among the best-supported learning techniques. Of ten techniques reviewed by Dunlosky et al. (2013), practice testing and distributed practice were rated of high utility because they benefit learners of different ages and abilities across many tasks, including in educational contexts; rereading was rated of low utility, and practice testing is
 55 recommended in its place. In a systematic review of 50 experiments in schools and classrooms, 57% of the effect sizes showed medium or large benefits of retrieval practice, across education levels, subjects, retention intervals and test formats, although only 6% of the experiments were conducted in non-WEIRD countries (Agarwal, Nunes and Blunt, 2021). Spacing the retrieval episodes adds to this: across 29 studies, spaced retrieval practice showed a strong benefit for retention over massed retrieval practice ($g = 0.74$), whereas expanding intervals were not superior to uniform ones (Latimier, Peyre and
 60 Ramus, 2021). Elaborative interrogation, by contrast, was rated of only moderate utility, because its benefits had not been adequately evaluated in educational contexts (Dunlosky et al., 2013).

AI tutors built on large language models add a specific risk to self-study: they can state things that no source supports. Ji et al. (2023) distinguish intrinsic hallucinations, which contradict the source content, from extrinsic hallucinations, which the source can neither support nor contradict. In a field experiment with nearly a thousand high school mathematics students, a
 65 GPT-4 tutor without safeguards gave a correct answer to the practice problems only 51% of the time, and students who had used it scored 17% lower than the control group once access was removed; a second tutor, whose prompt contained the teachers’ solutions and hints, largely mitigated this harm (Bastani et al., 2025). Kestin et al. (2025) likewise did not rely on GPT-4 alone but enriched the prompts with complete step-by-step solutions, and found that students learned significantly more in less time with this tutor than in an active learning class. Retrieval-augmented generation pursues the same aim by
 70 drawing relevant information from an external knowledge base into the answer, although mitigating hallucinations remains an open challenge there (Li et al., 2025). In all three approaches, the accuracy of the tutor is meant to rest on material supplied to it rather than on the model’s own knowledge.

Studies on audio-assisted reading compare reading while listening with reading alone. A meta-analysis of 30 studies found only a trivial overall benefit of reading while listening for comprehension ($g = 0.18$); the benefit was limited to studies in



75 which reading was paced by the experimenter ($g = 0.41$), and there was no reliable effect when reading was self-paced ($g = 0.06$) (Clinton-Lisell, 2023). Voice assistants and smart speakers have only recently begun to be used in schools and universities; their role in the classroom is still at an early stage of research, results so far are limited, incomplete and unorganized, and the lack of many languages is one of the challenges (Terzopoulos and Satratzemi, 2020).

Short and mobile learning have been described in their own right. A review of microlearning and social media cites
80 definitions ranging from lessons of 30 seconds to five minutes to any content that can be taken in within less than ten minutes (Denojean-Mairet et al., 2024). Mobile learning has been described as learning while travelling, driving, sitting or walking that lets learners use small amounts of time and space (Traxler, 2007).

Writing about one's own learning can support self-regulated learning, which Panadero (2017) describes as a core conceptual framework for understanding the cognitive, motivational and emotional aspects of learning. According to Nückles et
85 al. (2020), written text relieves the learner as an external representation and memory aid, but learning journals need instructional support if learners are to exploit this. In a study with 38 learners of English as a foreign language, writing reflective journals raised the students' awareness of their own strategies and may have contributed to their wider use (Zhang et al., 2022). A meta-analysis of 32 studies found large effects of AI-based interventions on self-regulated learning ($g = 1.613$) and self-directed learning ($g = 1.111$), moderate effects on cognitive and metacognitive regulation, and the strongest
90 effects in the forethought phase (Achuthan, 2025). In the journal studies, the learners write the journals themselves.

Narrative presentation has a measurable advantage for understanding and memory. A meta-analysis of over 75 samples with more than 33,000 participants found that stories were more easily understood and better recalled than expository texts, a result that was robust and not moderated by study characteristics (Mar et al., 2021). For science communication with non-expert audiences, Dahlstrom (2014) defines narratives by cause-and-effect relationships between events over a period of time
95 that affect particular characters, and associates them with increased comprehension, interest and engagement; because narratives are also intrinsically persuasive, their use raises ethical considerations.

Taken together, this work leaves a gap that the present paper addresses, in five respects. First, the AI tutors in the studies rely on solutions supplied by teachers or researchers, whereas here the learner supplies a fact document of his own, and every expected answer must occur word for word in it. Second, the work on narrative texts compares stories with expository texts;
100 none of the studies reviewed here examines a narrative version built from the same verified fact document, without losing or adding a fact, as a second way into the same material. Third, listening alone has been compared with reading in many studies, with no reliable overall difference in comprehension (Clinton-Lisell, 2022), and answering aloud during retrieval practice has been compared with typing, with few differences (Putnam and Roediger, 2013); none of the studies reviewed here, however, combines listening alone with spoken answers in self-study, as the voice sessions studied here do. Fourth, in
105 the journal studies the learners write the journals themselves, whereas here the AI assistant keeps the learning record on the learner's behalf. Fifth, the studies concern school classes, courses and samples of many learners; this paper describes the self-study of a single adult learner in the geosciences, who is also the developer of the learning environment. The first two respects correspond to RQ1, the third to RQ2 and the fourth to RQ3.



3 Materials and methods

110 3.1 Cases

The study draws on five cases: four learning projects, on the geology of southern Norway, Earth's geological eras, the geology of Germany and volcanology, and a narrative version of the fact document on the geological eras (Table 1, Fig. 1). Each case has a fact document and a learning path that passed the check script; the fact document on volcanology is in English, the others are in German. Sessions were run on two cases; the other three document how learning paths are built.

115 **Table 1. Cases, status 13 October 2026. Words: words of learning material in the learning path. Sessions run: sessions in the learning records; the first voice session on southern Norway, on 29 September 2026, is not counted because it went into a second record created by mistake (Sect. 4.7). Learning path: the latest version with a documented check run (Sect. 4.2); after version 10 of the fact document on the geological eras, both of its learning paths were cut anew, into 15 and 14 sessions, without a documented full check run.**

Case	Learning path	Words	Sessions in learning path	Questions	Sessions run	Review sessions
Southern Norway	version 7	14,267	31	93	15 (11 on version 6, 4 on version 7)	1
Geological eras, fact document	version 7	3,595	11	33	0	0
Geological eras, narrative version	version 8	4,692	10	30	10	1
Geology of Germany	version 1 (file name: version 6)	9,682	25	75	0	0
Volcanology, stage 1	version 4	19,519	38	114	0	0

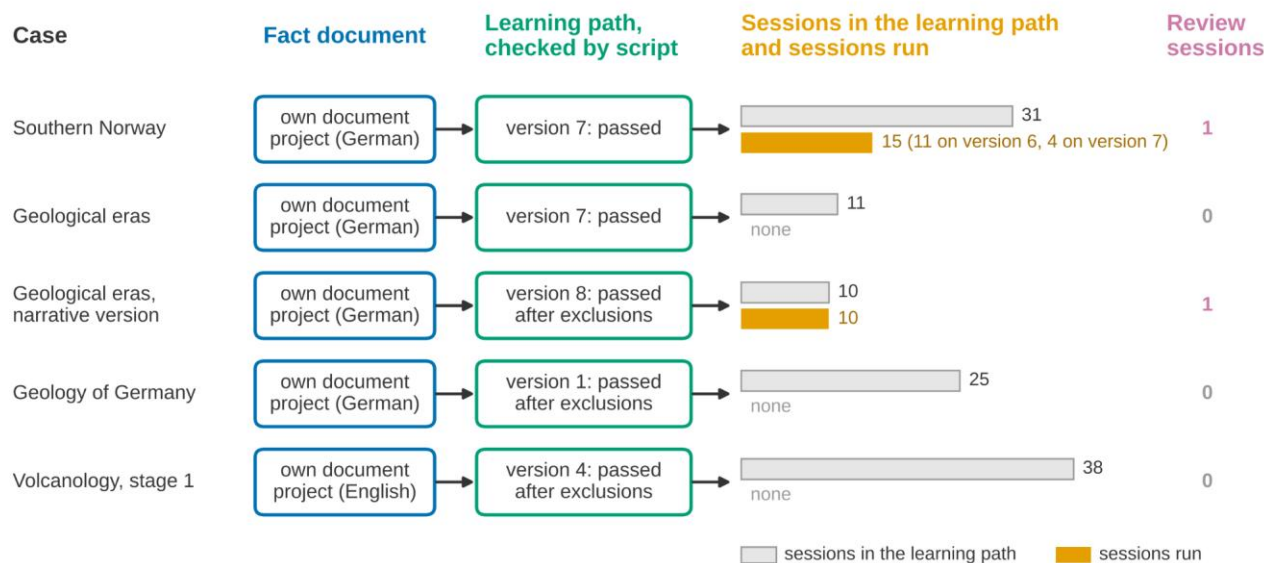


Fig. 1. The five cases at a glance, status 13 October 2026. Each fact document was built in a document project of its own, and each learning path passed the check script, three of them only after exclusions (Table 4). Sessions were run on two cases; for southern Norway, the sessions run belong to versions 6 and 7 of the learning path, and the bar of sessions in the learning path shows version 7 (Table 1).

3.2 Working cycle and interaction

The assistant was Claude (Anthropic), model Claude Opus 5.5 at the effort setting medium. Learning paths were built on the claud.ai website in a web browser, where the assistant can run code; the voice calls ran in the Claude app for Android. A skill is a folder of written instructions that the assistant loads when a request names it or matches its description; lernpfadsitzungen consists of a rule file, a check script with tests and an example pair of fact document and learning path (Appendix A). The format of the sessions was set on 27 September 2026: short lessons alternating with short quizzes; a Socratic part was rejected. From the start, the learner wanted to work through a learning path either on screen or interactively by voice, switching at any moment, and asked for a skill that could be called in any chat. As in *Talk, Check, Decide* (Mohr, 2026), he worked largely by voice; in the chats in which learning paths were built, he steered mainly with short instructions, in one of them with 22 of 41 inputs of three words or fewer. Rules for the assistant arose from failures and requests during sessions and were written into later versions of the skill (Sect. 4.7, Appendix A); the division of work between the two channels described below is one of them.

Building a learning path requires code execution and therefore does not run in voice mode; learning with a finished pair of fact document and learning path works in any mode. During a voice call, the assistant could neither update the learning record nor read the excerpt of a session from its file, and it declined to recite the excerpt from memory. The work was therefore divided between two channels: the text chat to prepare a session and to keep the learning record, the voice call to listen and answer (Fig. 2).



A session in voice mode runs in six steps:

1. The learner names the session, for example ‘session 5’ or ‘next’ (translated from German); the learning record is attached or lies in the project.
2. In the text chat, the assistant shows the excerpt and the three questions without their answers.
3. The learner switches to the voice call.
4. The assistant reads the excerpt in full and announces its end.
5. It asks the three questions one at a time, waits for each answer and grades it.
6. Back in the text chat, the assistant updates the learning record and hands it over as a file.

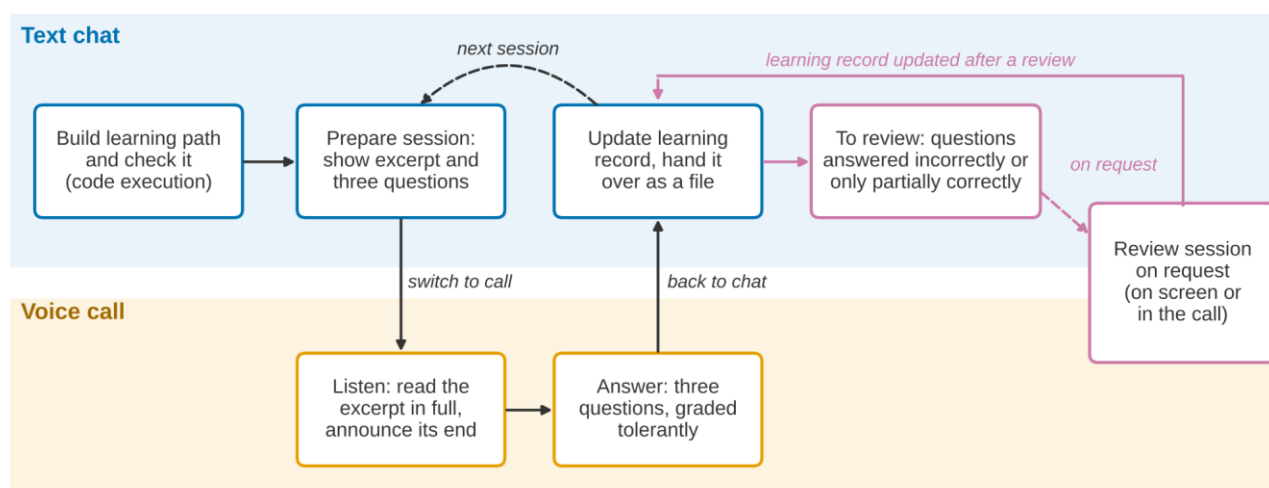


Fig. 2. Working cycle of a session in voice mode, as specified in version 2.6.6 of the skill (Sect. 3.6): the text chat for building, preparing and keeping the learning record, the voice call for listening and answering. Questions answered incorrectly or only partially correctly are collected in the list *To review* and taken up in review sessions on request, which run on screen or in the call; after each review, the learning record is updated (Sect. 3.5.3). Pink marks the review part; dashed arrows mark steps that follow only on request or with the next session.

3.3 Building a fact document

Each of the four source fact documents was built in a document project of its own before any learning path existed, with a workplan, a decision log and a record of all changes, and the narrative version was written within the project on the geological eras (Sect. 4.6); the skill lernpfad-sitzungen presupposes a finished fact document and never changes it; the fact documents were built with the skill template-dokumentprojekt. The approach goes back to *Talk, Check, Decide* (Mohr, 2026), in which the author had learned to build skills with which fact documents are written and checked. The learner set topic, level and sources for each case. For southern Norway, he asked for a fact document of first-class research at an advanced level, through which a chatbot would then guide him (translated from German). For Earth’s geological eras, all names, ranks and boundary ages were to be taken only from the International Chronostratigraphic Chart of the ICS, edition



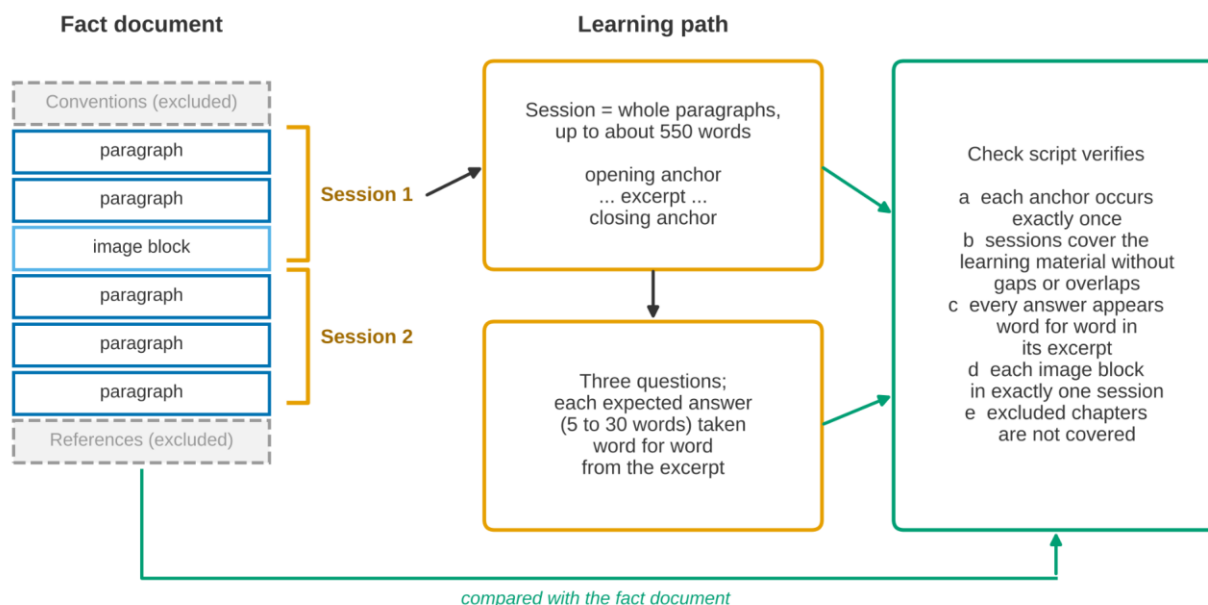
165 v2026/06, in the file he supplied, down to the rank of epoch; the questions were to ask not for decimal places but for order,
major boundaries and orders of magnitude. The fact document on the geology of Germany was based on open-access
sources, and the one on volcanology was taken word for word from openly licensed textbooks.
Someone who wants to use the skill with a fact document of their own needs a document in Markdown that meets the five
conditions in Sect. 3.4 and has been checked against its sources before a learning path is built, because the check script
170 compares the learning path only with the fact document (Sect. 4.1). Skills for writing and checking such documents are
described in *Talk, Check, Decide* (Mohr, 2026).

3.4 Building a learning path

A fact document can be cut into sessions when it meets five conditions (Fig. 3):

1. The learning material consists of whole paragraphs; a session never splits a paragraph.
- 175 2. Every paragraph that opens a session is at least six words long, because the check script forms anchors only from
six words upward.
3. In tables, the first cell of each row carries the name of the entry, not its rank or number, so that rows can serve as
anchors.
4. Chapters that are not learning material, such as conventions, tables of units and references, are named in the
180 learning path and passed to the check script as exclusions.
5. Figures stand in image blocks, which are not learning material but are assigned to exactly one session.

When the learning path has been cut, the check script compares it with the fact document and verifies the anchors, the
coverage of the learning material, the expected answers, the image blocks and the exclusions (Fig. 3; Sect. 4.2).



185 **Fig. 3.** How a fact document is cut into the sessions of a learning path and what the check script verifies when it compares the
 learning path with the fact document (Sect. 4.2); word limits as in Sect. 3.5.2. Blue boxes are paragraphs of learning material, the
 light blue box an image block and the grey dashed boxes excluded chapters; orange marks the sessions and questions of the
 learning path and green the check script. The paragraphs and chapters shown are schematic.

The learning path for southern Norway in version 7 already asked more about causes, processes and interpretations, at the
 190 learner's request; version 2.5.0 then made this the rule (Sect. 1). By the learner's account, a session took five to ten minutes,
 as intended.

3.5 Running a session

3.5.1 Listen: verbatim reading

In voice mode, only the form of the text may change, never its content:

- 195
- The excerpt is read in full in every call, even if it was shown in the chat before.
 - Source references are not read out; a source that is the subject of a sentence becomes 'the authors' or 'a source'.
 - Units, symbols and abbreviations are spoken in full, so that '425 Ma' is read as '425 million years'.
 - Decimal digits are spoken one by one, so that 2.43 is read as 'two point four three' (translated from German).
 - Image blocks are skipped, and the end of the excerpt is announced.
- 200
- The questions are asked as worded in the learning path, without hints, even when the learner asks for one; a question is rephrased only at the learner's request, in its form, not its content.



Figures and maps reach the learner only on screen: in screen mode, image blocks are shown with their captions, and for the narrative version a separate booklet of 15 maps served to show the cited maps and to answer follow-up questions.

3.5.2 Answer: questions and grading

205 Each session covers whole paragraphs, aimed at up to about 550 words of learning material (in the example pair, 28 of 30 sessions stay within this and the longest has 659 words), and ends with three questions, each with an expected answer of 5 to 30 words taken verbatim from the excerpt. This size follows from the intended length of a session, five to ten minutes, so that a session fits into a car journey or a short break.

Answers are graded tolerantly: the core of the answer counts. The learning records use three grades (Table 2). A fixed scale, 210 the same in speech and in writing, was proposed after a spoken ‘not quite right’ had been entered as ‘not correct’ in the record (translated from German).

Table 2. Grades with rules and examples.

Grade	Rule	Example
correct	the core of the answer is given; a value within the stated period counts as correct; missing details are noted next to the grade	the core is named, one detail is missing and noted
partially correct	only part of the core is given, or the order of magnitude is confused, with a note on the unit	Ma given instead of Ga
not correct	the core is missing or wrong, or no answer is given	no answer

When the tolerance rule was introduced, earlier answers were regraded. The learning record keeps the earlier grade and the date of the change next to the new grade instead of overwriting it. Speech recognition errors, such as ‘55 Ma’ heard as ‘55 215 times’ (Sect. 1) or ‘chapter 2 or module 2’ for session 2 (translated from German), were interpreted correctly by the assistant in the recorded cases.

3.5.3 Memorize: learning record and review sessions

Table 3 shows the structure of a learning record. The example entries are invented, because the records themselves are not published.

220 **Table 3. Structure of a learning record, with invented example entries.**

Part	Content	Invented example
Header	fact document and learning path, each with version and date; number of sessions; date of last update; grading rule	fact document FACTS_Example_Basin, version v3, 2026-01-10
Progress	one row per session: session, chapter, state, date, correct answers	4; 2 Basin fill; done; 2026-01-14; 2 of 3
Session blocks	per session: date, mode, opening and closing anchor, then each question with the learner’s answer, the grade and the verbatim answer	4.2; What filled the basin first?; river sand; partially correct (lake deposits missing); first lake deposits, then river sand
Reviews	per review: date, mode, topics, then each question with the learner’s answer, the grade and whether the entry is done	review 1; 2026-01-20; both; basin fill; 4.2; lake deposits, then river sand; correct; done
Open questions	questions left open, with session and date	4; Why did the lake dry out?; 2026-01-14



Part	Content	Invented example
<i>To review</i>	questions to be reviewed, with keyword and state	4.2; basin fill; open

At the learner’s request, questions answered incorrectly or only partially correctly are collected in the list *To review* and later taken up in review sessions. A review session starts only when the learner asks for one with ‘review session’ or ‘start review’ (translated from German); a plain ‘repeat’ is reserved in voice mode for hearing the last paragraph or question again. A review covers at most four topics, the oldest entries first. In the default mode, each topic begins with an introduction, continues with the core passage read verbatim from the fact document and a closing summary, and ends with the original question. Trial runs are not entered in the learning record. Besides this default mode, the learner can ask for a review that reads the passages only verbatim or one that only explains them. A learning record belongs to one pair of fact document and learning path; if a version differs, the assistant reports this instead of continuing silently. When southern Norway moved to a new learning path, eight new sessions whose material lay entirely in completed sessions were credited, and five with new material were closed at the learner’s request without being asked.

3.6 Data

The data come from four kinds of source. First, the chats in which the fact documents and learning paths were built and the sessions were held; inputs given by voice are available only as the text produced by the speech recognition. Second, the learning records, four files, one for each case except the fact document on the geological eras, for which none was created; the records on the geology of Germany and on volcanology contain no sessions yet, and the second record created by mistake for southern Norway (Sect. 4.5) is not counted. The assistant kept the records on the learner’s behalf after each session (Sect. 3.5.3); this paper reports totals per case, not the records themselves. Third, the output of the check script for every learning path in Table 4 (Sect. 3.4). Fourth, four method analyses of the chats in which the skill was developed and used, written between 1 and 9 October 2026; they are cited as a data basis and are not published.

Every statement in this paper rests either on a chat passage read in its original wording or on a file checked with a measured value. Statements known only from machine-generated chat summaries or from project notes were checked against one of these sources or not used.

Three limits follow from the way the data arose. The learning records were written by the assistant, not by the learner, and are only as accurate as the grading in the session. The method analyses are self-analyses: in each case the assistant analysed a chat in which it had itself taken part. Three of them were checked by a separate agent of the same model family; it found 18 deviations in one analysis, and in the other two it assessed 85 and 93 statements, of which 30 and 37 deviated, and reported 31 and 34 findings. The first analysis was examined only by a later meta-analysis. For one chat, the saved transcript begins after a context compaction, so that the wording of 74 of its 86 inputs rests only on the chat log the assistant kept during the work.

The study is a self-experiment in which the learner is also the developer of the learning environment. Methodologically, it comes closest to autoethnography, which describes and systematically analyses personal experience and in which reliability



refers to the narrator's credibility (Ellis, Adams and Bochner, 2011). Success is therefore judged by the learner's own account, not by a test, and on three counts: whether learning was possible at a time with hardly any free time, for example on the move or in short sections; whether the sessions were short and fitted the learner so well that five minutes were worth
255 spending on the next one; and whether the motivation to learn again grew because the material was presented in the way the learner learns best. The grades in the learning records (correct, partially correct, not correct) are process data. They show how a question was answered in a session and are not used as a measure of learning success.

The data cover the period from 27 September 2026, when the format of the learning sessions was set, to a cut-off date of 13 October 2026. All chats were held in German; in the case on volcanology, the fact document, the excerpts read aloud and the
260 questions were in English. Quotations from them are given in English translation and marked as translated from German; hesitations, filler words and broken-off phrases from voice input are left out of the translation.

The sessions ran under different versions of the skill. Sessions 1 to 3 on southern Norway, version 6, ran under version 2.3.1, and sessions 4 to 11 presumably under version 2.4.0, the version that was installed in their chat when version 2.5.0 was built there. Three times the skill was revised in the chat in which sessions had just been run: version 2.4.0 after sessions 1 to 3,
265 version 2.5.0 after session 11, and the first line of version 2.6.0 after the ten sessions of the narrative version; that line was never installed. This follows the iterative skill development described in *Talk, Check, Decide* (Mohr, 2026), in which every use is a test of the skill and what fails goes back into it; here the revision happened in the same chat, between two sessions. The four sessions on version 7 and all ten sessions of the narrative version ran under version 2.5.0, as far as the chats show; in sessions 8 to 10 of the narrative version, the questions were already shown in the text chat before the call, a rule first
270 written for the first line of version 2.6.0 and installed only with version 2.6.5. Sections 3.2 and 3.5 describe the current version 2.6.6 (Appendix A), which brings together the rules that arose from these sessions; no session in the data ran under it or under its predecessor 2.6.5.

3.7 Use of AI in the preparation of this paper

All work on this paper was done with Claude Opus 5.5 (Anthropic), in a web browser and in the Claude app for Android,
275 including the notes the author dictated as the basis for writing. Claude analysed the chats and learning records, collected the facts and checked them against their sources, ran the check script on the learning paths, searched and checked the literature, and drafted the English text from the author's German material and decisions. Claude also drafted the wording of the rules of the skill lernpfad-sitzungen and built its versions, which the author installed and used. The author determined the content, structure and conclusions.



280 4 Results

4.1 Building the fact documents

The check script compares a learning path only with its fact document (Sect. 4.2); errors that the fact document itself contains pass into the learning path unnoticed. How the fact documents were checked therefore matters. The fact document on volcanology was extracted word for word from textbook PDFs with a project script, which compared it word by word
 285 against the raw text and found no foreign words. Later steps nevertheless found two serious errors that this comparison could not detect, because it recognizes neither paragraph boundaries nor the kind of text: a chapter with three instead of eight paragraphs, and 14 words of a figure description in the running text. For the geology of Germany, the learner objected to photographing book pages, and a search for open-access sources followed. Each chapter then ran in a fixed cycle: the assistant requested up to six freely readable sources in one message, the learner supplied them, and the assistant cited them
 290 with page numbers and tried each wording template on a copy before insertion. Only the consistency check at the end found 49 first mentions without the English term in brackets and mixed British and American spelling, which the checks of the individual templates had not counted. The narrative version of the geological eras was built from an inventory of the 144 sentences of its source fact document (Sect. 4.6).

How long building took can be read from the chats; working hours were not recorded. The fact document on southern
 295 Norway and its first learning paths were built in one chat from 27 to 30 September 2026; version 6 of the path was checked at the start of a session chat on 30 September, and version 7 followed later in that chat. The fact document on the geological eras and its narrative version were built in one chat from 2 October, the narrative version dated 3 October, and its learning path was checked on 5 October. The fact document and learning path on the geology of Germany took one chat from 2 to 9 October, those on volcanology one chat on 8 and 9 October; each path was checked in a chat of its own on 9 October.

300 4.2 Building the learning paths

Table 4 lists seven check runs of six learning paths; versions 5 and 6 of the path on southern Norway are the same learning path under two names. In each run, the skill's check script compared the learning path with its fact document (Fig. 3): every session needs an opening and a closing anchor that occurs exactly once in the fact document, the sessions together must cover the learning material without gaps or overlaps, and every expected answer must appear word for word in the excerpt of
 305 its session; in addition, every image block must be assigned to exactly one session, and excluded chapters must not be covered. All six paths passed. Three of them passed only after chapters that the learning path itself marks as no learning material had been named as exclusions; for the geology of Germany, one opening anchor also had to be set by hand. By default, the check script excludes reference chapters; exclusions named in the learning path replace this default (Table 6). The last column lists these corrections.



310 **Table 4. Check runs of the learning paths. Anchors: anchors found exactly once, of all anchors. Words: words of learning material covered by the sessions. Answers: expected answers found word for word in the session excerpt, of all answers. Southern Norway, version 5, is the example pair of the skill and was re-measured on 13 October 2026 with the script of version 2.6.6; the original run had counted 13,859 words. Version 6 was checked on 30 September 2026 with the script then in use, which already counted 13,890 words.**

Learning path	Sessions	Anchors	Words	Answers	Result	Exclusions and corrections until passed
Southern Norway, version 5	30	60/60	13,890	90/90	passed	none recorded
Southern Norway, version 6	30	60/60	13,890	90/90	passed	none recorded; same learning path as version 5, renamed; checked on 30 September 2026, when the headers of both files still read version 5
Southern Norway, version 7	31	62/62	14,267	93/93	passed	none recorded
Geological eras (fact document), version 7	11	22/22	3,595	33/33	passed	none recorded
Geological eras (narrative version), version 8	10	20/20	4,692	30/30	passed	two chapters excluded from the learning material (tables of units, references)
Geology of Germany	25	50/50	9,682	75/75	passed	first run 577 findings; chapters the path itself marks as no learning material excluded; reference chapter added to the exclusions; opening anchor of session 20 set by hand
Volcanology, stage 1	38	76/76	19,519	114/114	passed	two chapters excluded (scope and conventions, references)

315 4.3 Listen: reading in voice mode

Listening produced requests of its own. In both cases with sessions, the learner asked for the excerpt to be read a second time, in part or in full: twice for southern Norway, where two sessions were also restarted on request, and twice for the narrative version. In the narrative version, the third session gave rise to the rule that the end of the excerpt be announced. Since none of the studies reviewed in Sect. 2 combines listening alone with spoken answers in self-study, these requests cannot be set against them; they are reported here as observations from a single learner.

325 4.4 Answer: questions and grading

Sessions were run on two cases (Table 5, Fig. 4). Of the 25 sessions in the learning records, 23 took place entirely in the voice call, one on screen and one first on screen and then in the call; the modes therefore allow no comparison, and none is drawn. Nor are the two learning paths on southern Norway compared: they cover partly different material, ask different questions and ran under different versions of the skill (Sect. 3.6). By 13 October 2026, 17 of the 31 sessions of version 7 on southern Norway were done: 8 credited from version 6, 5 closed at the learner's request without being asked and 4 asked; 14 were still open. All 10 sessions of the narrative version were asked.



Table 5. Answers per case and learning path, status 13 October 2026. Grades after the regrading of three answers of version 6 under the tolerance rule (Sect. 3.5.2).

Case	Learning path	Sessions	Questions	Correct	Partially correct	Not correct
Southern Norway	version 6	11	33	12	8	13
Southern Norway	version 7	4	12	1	5	6
Geological eras, narrative version	version 8	10	30	16	6	8

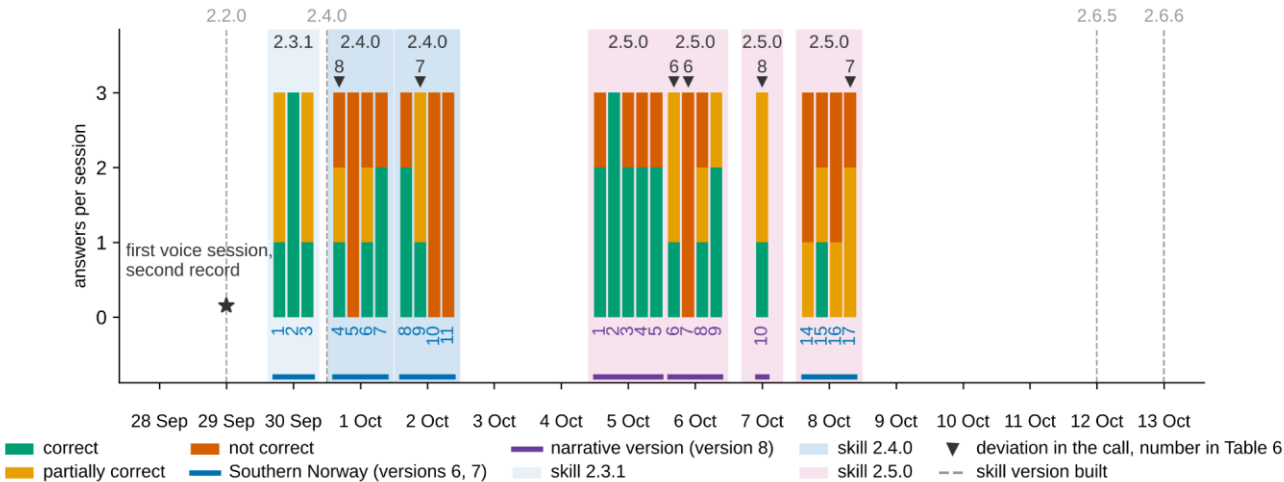


Fig. 4. Answers per session by date, for the two cases with sessions; numbers below the bars are session numbers. Shading marks the skill version in use, for sessions 4 to 11 on southern Norway inferred from the version installed in their chat; dashed lines the build dates of the versions dated to the day in Table A2, triangles deviations in the call with their number in Table 6, and the star the first voice session, which went into a second learning record. Process data, not a measure of learning success.

4.5 Memorize: learning records and review sessions

Each case with sessions is kept in one learning record, which the assistant updated after every session and review; for southern Norway, the first voice session of 29 September 2026 had gone into a second record created by mistake (Table 6). On 8 October 2026, the day of the last session on southern Norway, and unchanged until the cut-off date, the list *To review* held 35 entries for that case: 24 from version 6, of which two had been closed by regrading and one in a review session, and 11 from version 7; 32 were open. For the narrative version, it held 14 entries, of which 13 were open. Two review sessions were completed, one for each case, each with one topic.

4.6 Adapting the learning environment: the narrative version

The narrative version of the geological eras was written at the learner's request as a fact document of its own, so that nothing would be lost. Its basis was an inventory of the 144 sentences of the original fact document, and a second AI agent checked that every sentence was contained with its full content and that no fact had been added. The narrative version lies outside the skill lernpfad-sitzungen (Sect. 3.3); the skill then built a learning path of its own for it.



The learner's stated purpose was to approach the topic in a form that felt less dense with facts, as a way of thinking his way into it; the content was to be conveyed at a different level without rebuilding the fact document. The narrative version tells the geological eras as a journey through time from old to young. It was learned in ten voice sessions between 5 and 7 October 2026; of 30 questions, 16 were answered correctly, 6 partially correctly and 8 not correctly. No session was run on the fact document from which it was built, so this study cannot compare the two forms (Sect. 5).

4.7 Failures and the rules they produced

Table 6 lists the stumbling points documented in the learning records and in the analyses of the chats, with the case in which each occurred and the rule or remedy that followed. Fig. 5 places some of these rules in time, together with the building of the cases, the sessions and the versions of the skill.

Table 6. Stumbling points, their cases and the rules they produced.

No.	Stumbling point	Case	Trigger	Rule or remedy
1	Installed skill not found at the start	Southern Norway	the assistant searched its saved notes instead of the installed skills; in a second chat the start failed entirely	found after a hint from the learner; no rule recorded
2	Second learning record created	Southern Norway	no record file was attached, and the skill version in use provided for a new record in that case; earlier chats were not searched; session 1 ran twice	ask before creating a new learning record; in the skill since version 2.4.0
3	577 findings in the first check	Geology of Germany	the check script does not read back the chapters that the learning path marks as no learning material	name these chapters as exclusions; proposal that the check reads them from the learning path, not yet in version 2.6.6
4	Reference list taken as learning material	Geology of Germany	an own list of exclusions replaced the default, which excludes reference chapters	reference chapter added to the list; proposal that own exclusions supplement the default, not yet in version 2.6.6, where they still replace it
5	Call broken off before any answer	Geological eras, narrative version (review session)	no answers were available	learning record left unchanged; no result without evidence; rule in version 2.6.6
6	Questions asked other than those in the learning path	Geological eras, narrative version	in sessions 1 to 7 only the excerpt was shown before the call	excerpt and questions shown in the text chat before the call; in sessions 8 to 10 every question was asked as worded in the learning path; in the first line of version 2.6.0, and in the consolidated skill since version 2.6.5
7	Questions asked other than those in the learning path	Southern Norway	in session 9 of version 6 and session 17 of version 7, the call asked other questions than the learning path	the record notes the question actually asked; deviations are to be noted in the record since version 2.6.6
8	Content hint given in the call	Southern Norway; Geological eras, narrative version	the voice gave a hint on question 4.3 of version 6 on southern Norway, then graded correct, and on question 10.3 of the narrative version, then graded partially correct	no hint, even when asked; a hint is to be noted next to the grade since version 2.6.6
9	Image block assigned to no session	Volcanology	version 2.6.0, second line, skipped an image block that followed a heading	version 2.6.1

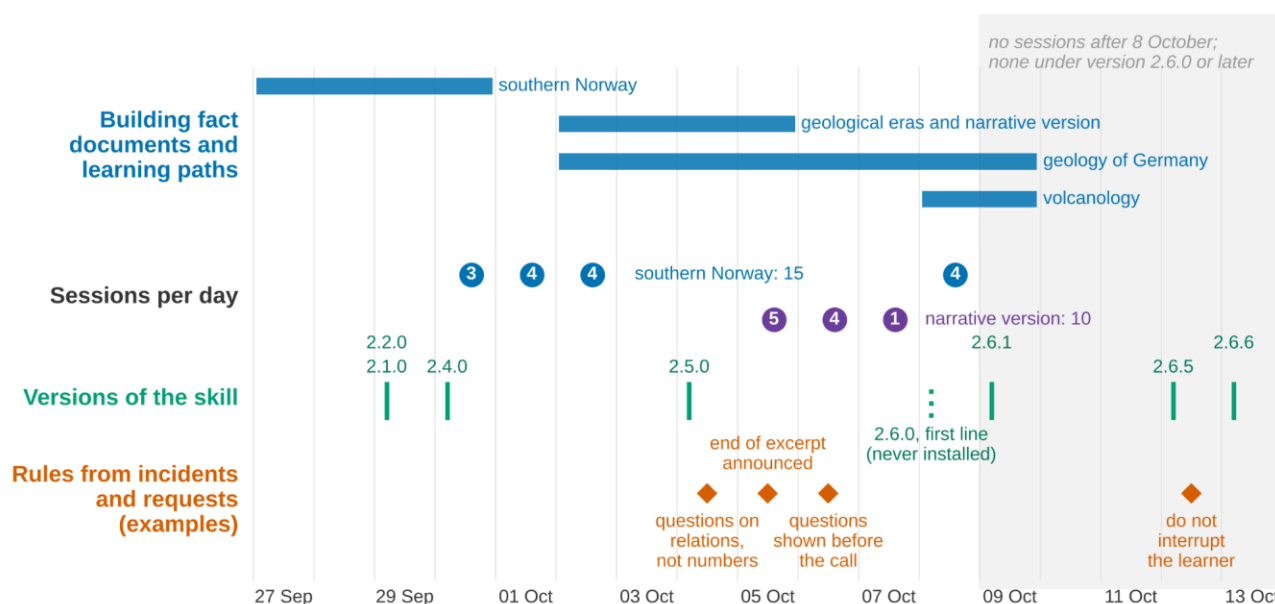


Fig. 5. Building and using the learning environment, 27 September to 13 October 2026: when the fact documents and learning paths were built (Sect. 4.1), how many sessions were run per day (Table 1), when versions of the skill appeared (Table A2; a version dated by a range is placed at its middle) and when some of the rules that arose from incidents and requests were set (Sect. 3.4, 4.3, Table 6). The first line of version 2.6.0 was never installed, and no session ran under version 2.6.0 or later (Sect. 3.6). Circles give the number of sessions per day in the colours of the cases in Fig. 4, diamonds mark the rules, and the dotted tick the version that was never installed.

4.8 The learner's assessment

By his own account, the learner met the three criteria set in Sect. 3.6 (Fig. 6). First, he could learn when he had little time. With about ten minutes in the car, the prepared session only had to be started and switched to voice mode on the phone, which ran hands-free in the car anyway; he listened and answered by voice. At work, he listened during breaks with headphones and spoke his answers in a quiet corner. 'For me, this was exactly the right way to learn' (translated from German). He used voice mode in almost all sessions and the screen in only three or four cases, because he learned mostly on the move; learning independently of place was one of his main aims, and the little time he had for learning was mostly time spent travelling. The learning records show a similar picture: 23 of the 25 sessions took place entirely in the voice call, one on screen and one partly on screen (Sect. 4.4). Second, he found the length of five to ten minutes exactly right. On car journeys of 30 minutes he often ran two sessions, and after the first he was motivated to run another, because after so short a session the results were quickly visible to him. Third, his motivation to learn again grew because the material was presented in the way he learns best. These are the learner's statements, not measured outcomes.

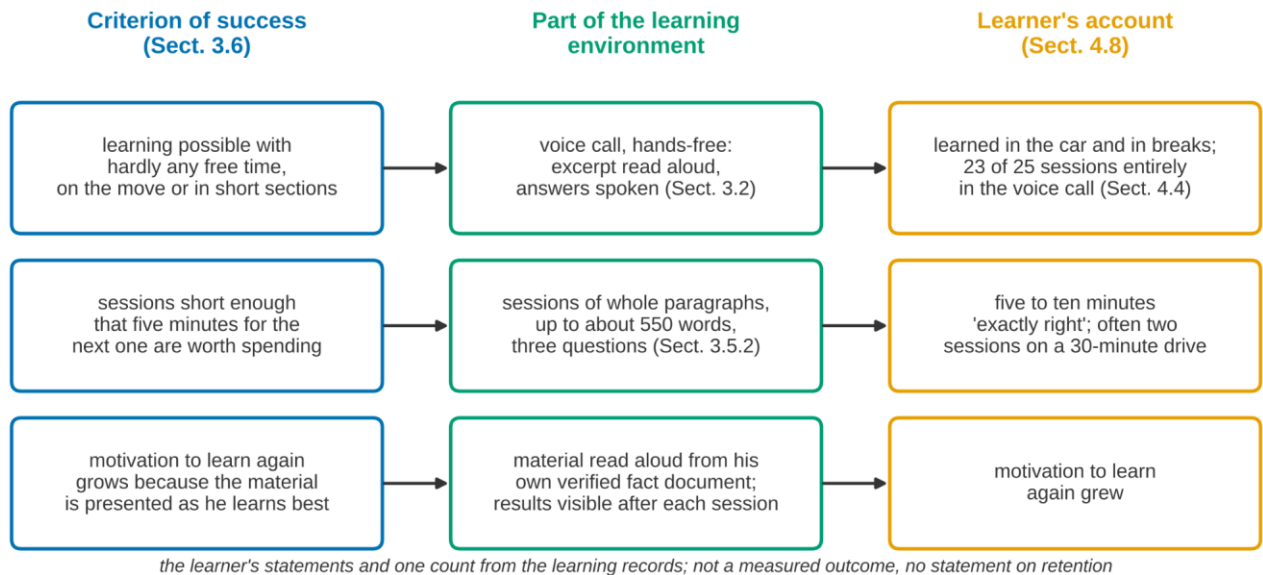


Fig. 6. The learner’s three criteria of success (Sect. 3.6), the parts of the learning environment that serve them, and the learner’s account of whether they were met (Sect. 4.8). The right-hand column reports the learner’s statements and one count from the learning records; it is not a measured outcome and says nothing about retention.

380 **5 Discussion**

The three research questions can be answered as far as the data of a single learner allow. RQ1: Each of the four source fact documents was built in a document project of its own, and a fifth, the narrative version, was written from one of them so that every one of the 144 sentences of its source was kept (Sect. 4.6); the skill cut each fact document into sessions whose expected answers the check script found word for word in the excerpt; all six learning paths passed, three of them only after chapters that are not learning material had been named as exclusions. The check binds the learning path to the fact document, not the fact document to its sources, so its correctness rests on how the fact document itself was checked, and there errors were found late (Sect. 4.1; Fig. 7). RQ2: In voice mode, the work was split between the text chat, which shows the excerpt and the questions and keeps the record, and the call, in which the excerpt is read in full and the three questions are asked and graded. The rules for the call came from incidents and requests in real sessions: questions shown before the call and asked as worded, decimal digits spoken one by one, the end of the excerpt announced and, in version 2.6.5, rules against interrupting the learner. RQ3: The learning record keeps every answer with its grade and collects incorrect and partially correct answers in the list *To review*, from which review sessions are built on request; with two completed reviews of one topic each, the data show how the record is kept, not what it achieves.

The verbatim principle works as a safeguard of the kind that the tutoring studies built into their prompts (Fig. 7). Bastani et al. (2025) and Kestin et al. (2025) achieved accuracy by supplying solutions to the model; here the fact document plays this



role, with the difference that the learner supplies it and every expected answer must occur word for word in the excerpt of the session. The design is meant to leave no room for either kind of hallucination distinguished by Ji et al. (2023), statements that contradict the source and statements that the source can neither support nor contradict. The principle does not enforce itself, however: in the first seven sessions of the narrative version, the assistant asked questions other than those in the learning path in two sessions, and only after the excerpt and the questions were shown in the text chat before each call were all questions asked as worded. Of the links from the sources to the grade, only the one from the fact document to the learning path is checked by the skill's check script; reading the excerpt, asking the questions and grading in the call rest on rules (Fig. 7). The same principle also sets a limit, because a question that the excerpt does not answer cannot be answered in the session.

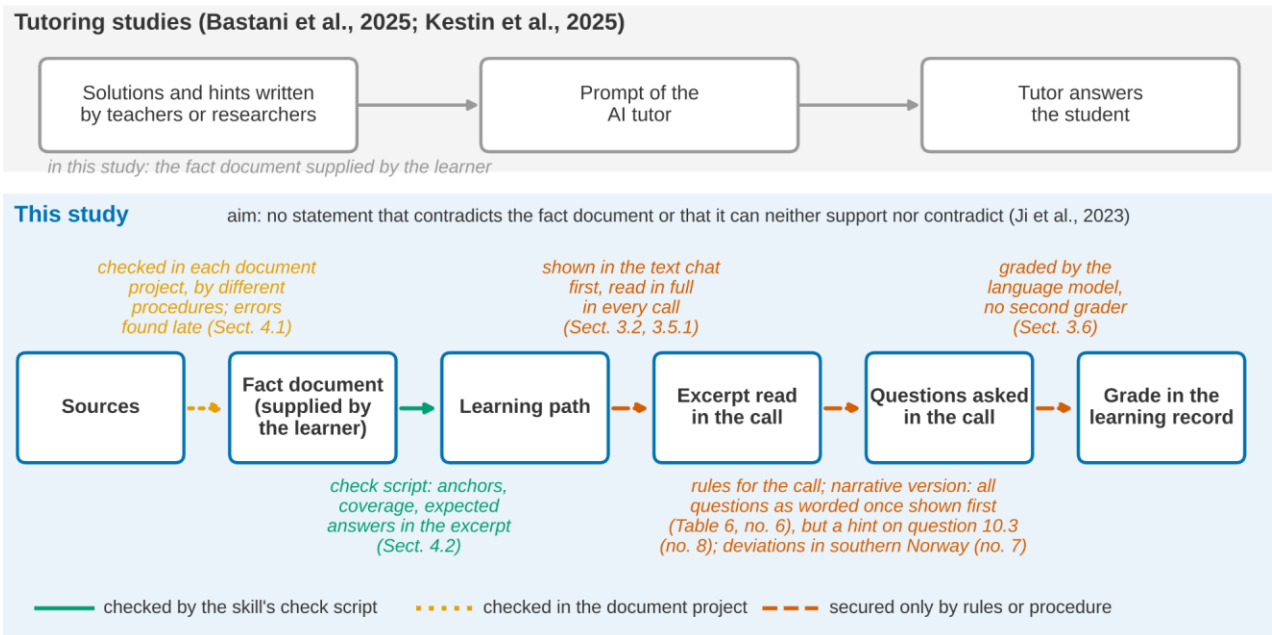


Fig. 7. The verbatim principle as a chain of safeguards, compared with the tutoring studies. In those studies, solutions and hints written by teachers or researchers bound the tutor through its prompt; in this study, the fact document supplied by the learner takes this role. Only the link from the fact document to the learning path is checked by the skill's check script (solid line). How the fact document matches its sources was checked in its document project, by different procedures (dotted line, Sect. 4.1). Reading the excerpt, asking the questions and grading in the call rest only on rules or procedure (dashed lines), and Table 6 shows that these rules did not always hold.

The narrative version shows that, in this study, a verified fact document could be given another form and still be bound to a checked learning path. The literature gives reason to expect an advantage of the narrative form: stories are understood and recalled better than expository texts (Mar et al., 2021), and narratives describe cause-and-effect relationships between events over time (Dahlstrom, 2014), as the narrative version does from one era to the next. Since no session was run on its source



fact document (Sect. 4.6), this study cannot test that expectation, and whether the advantage holds for a narrative version that must keep every sentence of a verified fact document remains open.

The sessions of five to ten minutes fall within the range given for microlearning, and learning in the car and during breaks matches the description of mobile learning (Sect. 2). The learner's account (Sect. 4.8) therefore concerns a familiar format, with the difference that each unit is cut from a fact document of his own and every expected answer stands word for word in it.

The grades in this study were given by the language model during the call. Without task-specific training, GPT-4 grades short free-text answers about as well as older hand-engineered systems but less well than specialized models, and its results can vary between runs (Kortemeyer, 2024). The grades in the learning records are therefore process data (Sect. 3.6); no second grader checked them (Fig. 7).

Open questions say little about learning in the present data. The rule is clear: when the learner asks something that the excerpt does not answer, the assistant does not answer from general knowledge but records the question as open and returns to the current quiz question. This happened once, in the first session in voice mode, when the learner asked for the time span of the Mesozoic. The question is not in the learning record kept for that case today, and the open-question lists of all four learning records are empty. The single case also showed two limits of the rule: the question was checked only against the excerpt, not against the whole fact document, and its wording was shortened in the record although it stood in quotation marks. An explanation mode in which general knowledge is allowed was later introduced for review sessions at the learner's request.

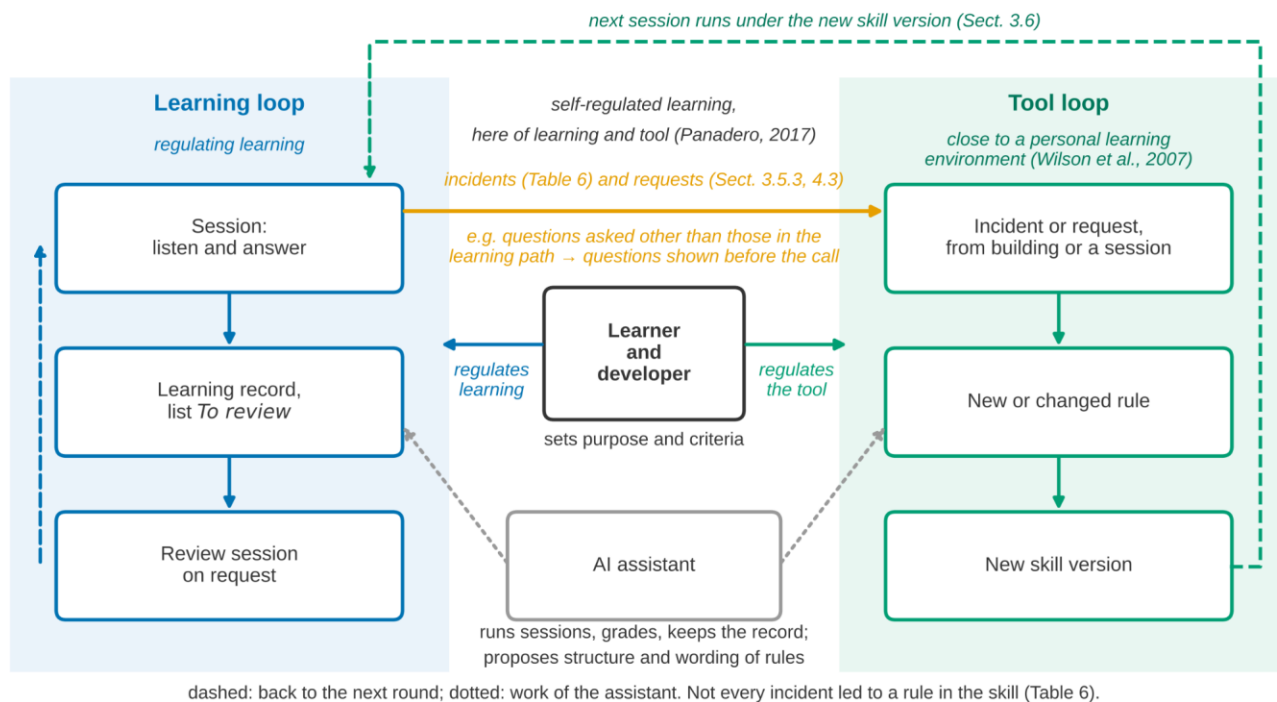
Review sessions in this study follow no schedule. They start when the learner asks for them and draw on a queue, oldest entries first, rather than on intervals planned in advance. The literature on spacing shows a strong benefit of spacing out retrieval practice over massing it, and no advantage of expanding over uniform intervals (Latimier, Peyre and Ramus, 2021); a queue served on demand allows spacing but does not ensure it. The present data cannot settle the question: only two review sessions were completed, each with one topic, and on 8 October 2026 the list *To review* held 32 open entries for southern Norway and 13 for the narrative version. A further review of the narrative version broke off before any answer was given. This paper therefore makes no claim about retention.

Directing the questions at relations and events comes close to what Dunlosky et al. (2013) call elaborative interrogation, a technique they rated of only moderate utility (Sect. 2). The present data do not show whether the change improved learning: the numbers of correct answers were not compared before and after version 2.5.0, and the questions changed together with the learning paths.

The learner in this study was also the developer of his learning environment, which he describes as a complete system of its own. The purpose and criteria of its parts came from him, for instance that the learning record should be portable and show progress first, while the assistant proposed their structure, such as binding the record to a version and collecting entries in lists. Several rules arose from his requests during sessions, among them the review sessions, their three modes and the use of general knowledge in the explanation mode. In terms of self-regulated learning (Panadero, 2017; Sect. 2), the learner



450 regulated not only his learning but also the tool he learned with (Fig. 8). This comes close to the idea of a personal learning environment, which is not a single program but the collection of tools a learner uses to meet his own needs, operating at a personal level (Wilson et al., 2007). A meta-analysis found large effects of AI-based interventions on self-regulated and self-directed learning (Achuthan, 2025); whether building the environment oneself adds to such effects cannot be judged from a single case.



455 **Fig. 8. The learner regulates two loops. In the learning loop, sessions fill the learning record and the list To review, from which review sessions are built on request. In the tool loop, incidents (Table 6) and requests from building learning paths and from sessions lead to new or changed rules and new skill versions; not every incident led to a rule in the skill. The next sessions run under the new skill version, which links the tool loop back to the learning loop. The learner set purpose and criteria; the AI assistant ran the sessions, graded the answers, kept the record and proposed the structure and wording of the rules (dotted lines). Dashed lines lead back to the next round.**

460

This study has five limitations. It rests on a single learner, who is also the developer of the learning environment. The data cover a short period, from 27 September 2026 to a cut-off date of 13 October 2026. There is no control condition, so effects of the method cannot be separated from those of the learner. Success is judged by the learner's own account. Sessions were run on only two of the five cases; the other three document how learning paths are built, not how they are used.

465

6 Conclusions and outlook

In this self-experiment, the learner assembled learning content from his own verified fact documents with an AI assistant and had it taught to him in more than one way: read aloud in a voice call, on screen, and as a narrative version of the same facts. He built the learning environment himself and adapted it while using it, instead of receiving a finished system that cannot be changed. By the criteria set in Sect. 3.6, he judges the experiment positively and reports that his motivation to learn again grew because the material was presented in the way he learns best; this is his own account, not a measured outcome. Two questions remain open: whether the sessions lead to lasting retention, which this study does not measure, and whether the approach serves other learners as it served its developer. Retention could be tested next by asking the questions of completed sessions again after fixed intervals and comparing questions that were taken up in review sessions with questions that were not; the list *To review* already records which entries were reviewed and when.

The method is not tied to the geosciences: the rules of the skill contain no subject content, and the skill works with any fact document that can be checked. In this study it was used for four geoscience topics, with fact documents in German and in English. In the author's assessment, it should also suit other subjects that rely heavily on factual and conceptual knowledge, such as biology, history or law, and less so subjects that are learned mainly by calculating or practising; this has not been tested.

Other learners need an account with which the assistant can run code, to build learning paths, and that offers a voice mode, to learn by ear; what is spoken in the call and the learning record are processed by the provider of the assistant.

Appendix A: The skill

Table A1 lists the commands of the skill in version 2.6.6, Table A2 its versions and Table A3 one session of its example pair. Commands are spoken or typed in German; the table gives an English gloss.

Table A1. Commands of lernpfad-sitzungen 2.6.6.

Command	Gloss	Effect	Screen	Voice
'Lernpfad erstellen', 'baue einen Lernpfad'	build a learning path, build me a learning path	cuts the fact document into sessions and checks the pair; needs code execution	yes	no
'Sitzung n', 'weiter'	session n, next	starts a session; 'next' takes the session after the last completed one	yes	yes
'Bildschirm', 'Sprache'	screen, voice	sets the mode; switching is possible at any time	yes	yes
'ich fahre', 'unterwegs'	I am driving, on the move	sets voice mode	yes	yes
'Pause'	pause	waits until 'weiter'	no	yes
'wiederholen'	repeat	repeats the last paragraph or question	no	yes
'langsamer'	slower	slows down	no	yes
'überspringen'	skip	moves on to the next question	no	yes



Command	Gloss	Effect	Screen	Voice
‘Stopp’	stop	ends the session and gives the record line	no	yes
‘Wiederholung ssitzung’, ‘Wiederholung starten’	review session, start review	starts a review session	yes	yes
‘wörtlich’, ‘erklären’, ‘erkläre mir den Inhalt nochmal’, ‘beides’	verbatim, explain, explain the content to me again, both	sets the mode of a review; ‘beides’ is the default	yes	yes

A plain ‘wiederholen’ does not start a review session, because in voice mode it repeats the last paragraph or question. By the author’s account, the commands also work when given in English, because the language model interprets them in either language; the example pair, by contrast, is in German (Code and data availability).

490 **Table A2. Versions of lernpfad-sitzungen. Earlier versions were specific to southern Norway; the number 2.6.0 was assigned twice, to two independent lines of development. Version 2.6.5 was built from 2.6.1 and took over the rules of the first line in new wording, except rephrasing on request, which followed in 2.6.6.**

Version	Date	Change
2.1.0	by 29 September 2026	function ‘build a learning path’ with the check script; sessions of whole paragraphs up to about 550 words, three questions with verbatim answers
2.2.0	29 September 2026	learning record file
2.3.1	by 30 September 2026	revision; its order relative to 2.2.0 could not be determined
2.4.0	30 September 2026	review sessions in the modes verbatim and explain; more tolerant grading; ask before creating a new learning record
2.5.0	3 to 5 October 2026	questions on relations and events; version line in the learning path; questions asked as worded, without hints, in voice mode
2.6.0, first line	rules proposed from 5 October, package 7 to 8 October 2026	questions shown before the call, decimal digits spoken one by one, end of the excerpt announced, ‘both’ as default review mode
2.6.0, second line	9 October 2026	image blocks; learning objectives guide the questions
2.6.1	9 October 2026	image block after a heading assigned to the following session
2.6.5	12 October 2026	2.6.1 with the rules of the first line of 2.6.0 in new wording, except rephrasing on request; rules against interrupting the learner
2.6.6	13 October 2026	a question may be rephrased on request, without a hint; no hint even when asked; counts given during the call are provisional; deviations and hints are noted in the learning record; nothing is entered for a call broken off before any answer; licence file

Table A3. Session 3 of the example pair, learning path version 5, translated from German; source references are left out.

Part	Content
Section	2.1
Opening anchor	paragraph beginning ‘The Precambrian basement of the area belongs’
Closing anchor	paragraph ending ‘55 Ma’
Length	496 words of learning material



Part	Content
Excerpt, first sentence	The Precambrian basement of the area belongs to the Fennoscandian Shield, also called the Baltic Shield.
Question 3.1	How long is the Transscandinavian Igneous Belt according to the two sources cited?
Expected answer 3.1	a chain of batholiths about 1400 km long from southeast Sweden to northwest Norway; a second source gives 1500 km from Skåne to Lofoten
Question 3.2	Where do the Caledonian allochthons come from according to current understanding?
Expected answer 3.2	some of these allochthons from Baltica, others probably of exotic origin, partly from the margin of Laurentia, which collided with Baltica
Question 3.3	Which rift phases are distinguished for the Froan Basin and the Norwegian Sea?
Expected answer 3.3	Late Permian to Early Triassic, Middle Jurassic to earliest Cretaceous, and middle Campanian until break-up with seafloor spreading at about 55 Ma

Code and data availability

495 The skill lernpfad-sitzungen, version 2.6.6, is available under the MIT License at <https://github.com/nordie69/Listen-Answer-Memorize>. The package, a .skill file that is installed as a skill in Claude, contains the rule file SKILL.md, the check script lernpfad.py with its tests, the licence file and an example pair: a fact document in German on the geology of southern Norway and its learning path in version 5 with 30 sessions, which can be reused under the same licence. The summary figures per case on which this paper is based are reported in the paper itself; the complete learning records and the method
 500 analyses of the chats are not published.

Author contributions

SM designed the study, built the fact documents, the learning paths and the skill lernpfad-sitzungen with the AI assistance described in Sect. 3.7, ran all learning sessions and prepared the manuscript.

Competing interests

505 The author has declared that there are no competing interests.

Financial support

This research received no external funding.

References

Achuthan, K. (2025) ‘Artificial intelligence and learner autonomy: a meta-analysis of self-regulated and self-directed
 510 learning’, *Frontiers in Education*, 10, 1738751. Available at: <https://doi.org/10.3389/feduc.2025.1738751>



- Agarwal, P.K., Nunes, L.D. and Blunt, J.R. (2021) ‘Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms’, *Educational Psychology Review*, 33(4), pp. 1409-1453. Available at: <https://doi.org/10.1007/s10648-021-09595-9>
- 515 Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö. and Mariman, R. (2025) ‘Generative AI without guardrails can harm learning: evidence from high school mathematics’, *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. Available at: <https://doi.org/10.1073/pnas.2422633122>
- Cheek, K.A., LaDue, N.D. and Shipley, T.F. (2017) ‘Learning about spatial and temporal scale: current research, psychological processes, and classroom implications’, *Journal of Geoscience Education*, 65(4), pp. 455-472. Available at: <https://doi.org/10.5408/16-213.1>
- 520 Clinton-Lisell, V. (2022) ‘Listening ears or reading eyes: a meta-analysis of reading and listening comprehension comparisons’, *Review of Educational Research*, 92(4), pp. 543-582. Available at: <https://doi.org/10.3102/00346543211060871>
- Clinton-Lisell, V. (2023) ‘Does reading while listening to text improve comprehension compared to reading only? A systematic review and meta-analysis’, *Educational Research: Theory and Practice*, 34(3), pp. 133-155. Available at: <https://files.eric.ed.gov/fulltext/EJ1403866.pdf> (Accessed: 12 October 2026).
- 525 Dahlstrom, M.F. (2014) ‘Using narratives and storytelling to communicate science with nonexpert audiences’, *Proceedings of the National Academy of Sciences*, 111(Suppl. 4), pp. 13614-13620. Available at: <https://doi.org/10.1073/pnas.1320645111>
- Denojean-Mairet, M., López-Pernas, S., Agbo, F.J. and Tedre, M. (2024) ‘A literature review on the integration of microlearning and social media’, *Smart Learning Environments*, 11(1), 46. Available at: <https://doi.org/10.1186/s40561-024-00334-5>
- 530 Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J. and Willingham, D.T. (2013) ‘Improving students’ learning with effective learning techniques: promising directions from cognitive and educational psychology’, *Psychological Science in the Public Interest*, 14(1), pp. 4-58. Available at: <https://doi.org/10.1177/1529100612453266>
- 535 Ellis, C., Adams, T.E. and Bochner, A.P. (2011) ‘Autoethnography: an overview’, *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), 10. Available at: <https://doi.org/10.17169/fqs-12.1.1589>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A. and Fung, P. (2023) ‘Survey of hallucination in natural language generation’, *ACM Computing Surveys*, 55(12), 248. Available at: <https://doi.org/10.1145/3571730>
- 540 Kestin, G., Miller, K., Klales, A., Milbourne, T. and Ponti, G. (2025) ‘AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting’, *Scientific Reports*, 15(1), 17458. Available at: <https://doi.org/10.1038/s41598-025-97652-6>
- Kortemeyer, G. (2024) ‘Performance of the pre-trained large language model GPT-4 on automated short answer grading’, *Discover Artificial Intelligence*, 4(1), 47. Available at: <https://doi.org/10.1007/s44163-024-00147-y>



- 545 Landy, D., Silbert, N. and Goldin, A. (2013) ‘Estimating large numbers’, *Cognitive Science*, 37(5), pp. 775-799. Available at: <https://doi.org/10.1111/cogs.12028>
- Latimier, A., Peyre, H. and Ramus, F. (2021) ‘A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention’, *Educational Psychology Review*, 33, pp. 959-987. Available at: <https://doi.org/10.1007/s10648-020-09572-8>
- 550 Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H. and Wang, F.L. (2025) ‘Retrieval-augmented generation for educational application: a systematic survey’, *Computers and Education: Artificial Intelligence*, 8, 100417. Available at: <https://doi.org/10.1016/j.caeai.2025.100417>
- Mar, R.A., Li, J., Nguyen, A.T.P. and Ta, C.P. (2021) ‘Memory and comprehension of narrative versus expository texts: a meta-analysis’, *Psychonomic Bulletin & Review*, 28(3), pp. 732-749. Available at: [https://doi.org/10.3758/s13423-020-](https://doi.org/10.3758/s13423-020-01853-1)
- 555 01853-1
- Mohr, S. (2026) ‘Talk, Check, Decide: a skill-based workflow for writing and verifying geological excursion reports with an AI assistant’, *EGUsphere* [Preprint]. Available at: <https://doi.org/10.5194/egusphere-2026-5952>
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J. and Renkl, A. (2020) ‘The self-regulation-view in writing-to-learn: using journal writing to optimize cognitive load in self-regulated learning’, *Educational Psychology Review*, 32, pp. 1089-
- 560 1126. Available at: <https://doi.org/10.1007/s10648-020-09541-1>
- Panadero, E. (2017) ‘A review of self-regulated learning: six models and four directions for research’, *Frontiers in Psychology*, 8, 422. Available at: <https://doi.org/10.3389/fpsyg.2017.00422>
- Putnam, A.L. and Roediger, H.L. (2013) ‘Does response mode affect amount recalled or the magnitude of the testing effect?’, *Memory & Cognition*, 41(1), pp. 36-48. Available at: <https://doi.org/10.3758/s13421-012-0245-x>
- 565 Terzopoulos, G. and Satratzemi, M. (2020) ‘Voice assistants and smart speakers in everyday life and in education’, *Informatics in Education*, 19(3), pp. 473-490. Available at: <https://doi.org/10.15388/infedu.2020.21>
- Traxler, J. (2007) ‘Defining, discussing and evaluating mobile learning: the moving finger writes and having writ . . .’, *International Review of Research in Open and Distance Learning*, 8(2). Available at: <https://doi.org/10.19173/irrodl.v8i2.346>
- 570 Wilson, S., Liber, O., Johnson, M., Beauvoir, P., Sharples, P. and Milligan, C. (2007) ‘Personal learning environments: challenging the dominant design of educational systems’, *Journal of e-Learning and Knowledge Society*, 3(2), pp. 27-38. Available at: https://je-lks.org/ojs/index.php/Je-LKS_EN/article/download/29/3 (Accessed: 13 October 2026).
- Zhang, Y., Li, M., Chieh, Y. and Han, S. (2022) ‘On a path to becoming more self-regulated: reflective journals’ impact on Chinese English as a foreign language students’ self-regulated writing strategy use’, *Frontiers in Psychology*, 13, 1042031.
- 575 Available at: <https://doi.org/10.3389/fpsyg.2022.1042031>