Micro-bottleneck Internalization Protocol: An Operational Framework for Embedding Retrieval Practice, Spaced Repetition, and Self-Explanation into Daily Learning Workflows

Authors: Wenbo Feng, Wenbo Feng

Date: 2026-03-03T15:31:09+00:00

Abstract

[Objective] During the internalization phase (following the completion of preliminary understanding), this study aims to provide learners with a set of repeatable learning protocols, enabling retrieval practice, spaced repetition, and self-explanation to be embedded into daily routines with low friction while continuously targeting individual weak points. [Methods] Methodologically, the acquisition phase is distinguished from the internalization phase. Using "cue-response mapping" as the minimal unit of analysis for internalization training, the Micro-Bottleneck Internalization Protocol (MBIP) is proposed: the i-Zone is used to locate microscopic sticking points, and the Minimal Scenario (MS) is used to encapsulate reproducible training units. Exercises are organized according to the sequence of "MS—Retrieval First—Scenario-Based Self-Explanation (SBE)—Repeated Repetition (integrable with spaced scheduling)." Furthermore, a dual-track consistency mapping—comprising a "micro-reinforcement track" and a "macro-structure track" —along with failure backtracking rules, is employed to constrain the coupling between fragmented training and the overall structure. [Results] The study presents the operational rules, minimum functional criteria, and failure diversion logic of the protocol. It also provides recordable process performance indicators (such as initiation latency, frequency of sticking points, prompt dependency, and structural backtracking stability) and falsifiable predictions to align with subsequent empirical testing. [Limitations] This paper is a framework-oriented work and does not report quantitative results from reviews or effect sizes from controlled experiments; the applicability of the protocol is constrained by task decomposability and execution friction of the medium, and certain problems or knowledge points may be difficult to stably segment into MS. [Conclusion] MBIP systematizes the training objects and scheduling rules of the internalization phase, aiming to improve internalization efficiency per unit of time and providing operational constraints for learners' self-scheduled practice and the design of low-friction learning tools.

Full Text

Preamble

Micro-Bottleneck Internalization Protocol: An Operational Framework for Embedding Retrieval Practice, Spaced Repetition, and Self-Explanation into Daily Learning Workflows

The Chinese University of Hong Kong, Hong Kong

[**Objective**] This study aims to provide learners with a repeatable learning protocol for the internalization phase (following initial comprehension). The protocol is designed to embed retrieval practice, spaced repetition, and self-explanation into daily routines with low friction, while continuously targeting an individual' s specific weak points.

[**Methods**] Methodologically, this paper distinguishes the acquisition phase from the internalization phase. Using "cue–response mapping" as the minimal unit of analysis for internalization training, we propose the Micro-Bottleneck Internalization Protocol (MBIP). The protocol utilizes the "i-Zone" to locate microscopic "stalling points" and encapsulates reproducible training units within a "Minimum Scene" (MS). Practice is organized according to the sequence: "MS → Retrieval-first → Scene-Based Self-Explanation (SBE) → Repeated Reinstatement (integratable with spacing schedules)." Furthermore, a dual-track consistency mapping—comprising a "Micro-Reinforcement Track" and a "Macro-Structural Track" —along with failure-backtracking rules, ensures the coupling of fragmented training with the overall knowledge structure.

[**Results**] We present the operational rules, minimum functional criteria, and failure-triage logic of the protocol. Additionally, we provide recordable process performance indicators (such as initiation latency, stall frequency, prompt dependency, and structural backtracking stability) and falsifiable predictions to align with subsequent empirical testing.

[**Limitations**] As a framework-oriented work, this paper does not report quantitative review results or controlled experimental effects. The applicability of the protocol is constrained by task decomposability and execution friction; certain problems or knowledge points may be difficult to segment stably into Minimum Scenes.

[**Conclusion**] MBIP systematizes the training objects and scheduling rules of the internalization phase. It aims to improve internalization efficiency per unit of time and provides operational constraints for learners' self-scheduling practices and the design of low-friction learning tools.

Keywords: Internalization; Retrieval Practice; Spaced Repetition; Scene-Based Self-Explanation; Minimum Scene; Micro-Bottleneck; Learning Protocol; Self-Scheduling

Micro-Bottleneck Internalization Protocol (MBIP): A Dual-Track, Closed-Loop Framework for Everyday Internalization Practice

Feng Wenbo The Chinese University of Hong Kong, Hong Kong SAR, China

Abstract: [**Objective**] To provide a runnable protocol for the internalization phase that embeds multiple learning mechanisms into everyday routines while continuously targeting an individual' s weakest links.

[**Methods**] We distinguish the acquisition phase from the internalization phase and operationalize internalization with cue–response mapping. We propose the Micro-Bottleneck Internalization Protocol (MBIP): i-Zone identifies recurrent micro-stalls; Minimum Scene (MS) encapsulates reproducible training units; practice follows an MS → retrieval-first → Scene-Based Self-Explanation (SBE) → repeated reinstatement (optionally with spacing) chain; and a dual-track system (Micro-Track vs. Macro-Track) ensures structural integrity.

Macro-Track) with consistent mapping and failure backtracking constrains long-term coupling between fragment-level training and global structure. [Results] We specify operational rules, minimal functional criteria, and failure-routing logic, and propose recordable process indicators (e.g., initiation latency, stall frequency, prompt dependence, and stability of structural backtracking) together with falsifiable predictions to align future empirical tests. [Limitations] This is a framework paper and reports no quantitative review outcomes or controlled effect sizes. Applicability depends on task decomposability and implementation friction; a small portion of items may not be stably segmentable into MS units. [Conclusions] MBIP systematizes the training targets and scheduling rules of internalization to improve time efficiency and to provide actionable constraints for learner self-scheduling and low-friction learning tool design.

Keywords: Internalization; Retrieval practice; Spaced repetition; Scene-based self-explanation; Minimum scene; Micro-bottleneck; Learning protocol; Self-scheduling

1.1 背景问题——学习方法的边界与研究对象

Before discussing the effectiveness of learning methods, it is necessary to define the types of tasks they address and their applicable boundaries. Without distinguishing between the different tasks involved in the learning process, discussions regarding "universal learning methods" are prone to over-generalization and scope misalignment. From a process-oriented perspective, learning comprises at least two distinct categories of tasks: initial understanding during the acquisition phase (moving from inability to ability) and the long-term retention and enhanced accessibility during the internalization phase (the ability to stably extract and flexibly apply knowledge in the future). This paper distinguishes between these two at a methodological level [?, ?, ?, ?]. It should be noted that while these two stages may overlap and recur in real-world learning, this paper maintains an operational distinction for the purposes of conceptual clarification and the delimitation of the research object.

This paper does not presuppose the existence of a single strategy that is equally effective for all content during the acquisition phase. Different materials vary significantly in terms of cognitive demands and required external scaffolding—for example, understanding mathematical axioms, memorizing phone numbers, mastering programming syntax, or grasping abstract concepts. Consequently, this study limits its research scope to the universal process constraints of the internalization phase. These constraints describe "how to organize retrieval, practice, and reflection" without promising equivalent effect sizes across all tasks [?, ?, ?, ?]. More specifically, this paper focuses on micro-processes that can be formulated as "cue-response path mappings"and whose initiation and completion quality can be evaluated within a defined time window. For task types where response paths are difficult to encapsulate stably (such as highly open-ended writing), we discuss only their boundaries and variants without making strong commitments to specific outcomes.

1.2 学习过程的阶段划分与相关研究

To avoid conceptual confusion, this paper operationally divides the learning process into an acquisition phase and an internalization phase [?, ?, ?, ?]. This classification is functionally similar to several classical distinctions; subsequent sections of this paper will utilize comparative citations to clarify specific similarities and differences.

The Acquisition Phase refers to the process by which a learner first encounters new information and forms a usable representation of it. In this study, the minimum operational marker for this phase is defined as the learner' s ability to "independently complete basic comprehension tasks" [?, ?].

The Internalization Phase is the stage following initial comprehension, characterized by the enhancement of subsequent retrieval and transferability through retrieval practice, exercise, and reflection [?, ?, ?].

During the acquisition phase, strategies tend to rely more heavily on the type of material and external scaffolding. Teacher guidance and adaptive tools (including generative AI) can lower the barrier to entry by providing explanations, examples, and feedback; this paper makes no prior assumptions regarding the magnitude of these effects.

Despite the abundance of research on the underlying mechanisms, a gap remains at the operational level regarding the internalization phase: specifically, how learners can organize retrieval, practice, and reflection in a repeatable, low-cost manner during daily study, while consistently targeting their training toward unstable micro-components of their knowledge. While mechanisms such as retrieval practice, spaced repetition, self-explanation, and cognitive load control have been extensively studied, practical implementation remains a challenge.

讨论 [5][6][8][9][17]；本文关注的问题是把这些机制组织成可执行的组合协议，并让训练稳定

Pointing toward the location of individual bottlenecks [**?**]. In natural learning contexts, learners often use error logs (mistake notebooks) or peer explanations to advance their studies. However, during the internalization phase, these practices frequently exhibit observable failure signals that expose gaps in the "operational-level protocol." While error logs preserve incorrect examples and error-prone points, they typically use the "entire problem" as the storage unit. A single problem often contains multiple knowledge points and procedural steps, whereas a learner's actual weakness may only reside within a small segment. If review sessions focus primarily on replaying the entire problem, the training becomes diluted by a large number of already mastered steps, leading to increased review time and poorly defined bottleneck localization. Failure signals include repeatedly getting stuck at the same micro-step during practice, while the error log entry fails to compress the triggering cues and subsequent response paths into the minimal fragments required for direct training. Similarly, conventional notes often mix cues and explanations in the same space; during review, this tends to lead to passive re-reading rather than active retrieval, making it difficult to consistently expose and train specific micro-step blockages.

Explaining concepts to others helps organize narratives and generate explanations. However, peer explanation usually requires providing prerequisite knowledge and context for the audience, making it difficult to focus solely on one's own unstable local segments. Consequently, the content of the explanation is often forced into a complete start-to-finish narrative, where training objectives are driven by macro-level coherence rather than the individual's micro-level sticking points. Furthermore, explaining depends on the availability of others, time, and specific settings, making it difficult to establish high-frequency training that can be reproduced at any time. Failure signals include being able to explain the general idea clearly while remaining unable to initiate the next response at the same critical step during independent problem-solving, derivation, or coding.

1.3 研究问题、核心假设与目标

1. Research Questions and Working Hypotheses

Research Question (RQ): After a learner has achieved a preliminary understanding of the material, is it possible to construct a repeatable internalization protocol that stabilizes the focus of learning resources on an individual's micro-bottlenecks, thereby promoting the automation and transfer of response paths [**?**, **?**, **?**]?

Working Hypothesis (H): When internalization training is organized according to the chain of "Minimal Scenario (MS)–Active Retrieval–Scene-Based Self-Explanation (SBE)–Spaced Repetition," and satisfies the consistent mapping requirements of a dual-track learning system, it is theoretically more likely to demonstrate faster improvements in process performance per unit of time (e.g., smoother initiation of tasks and decreased dependency on prompts). The magnitude of this improvement is moderated by task type, material structure, and the learner' s prior knowledge level.

Micro-analytical Units and Operational Definitions: This paper operationalizes the smallest trainable unit during the internalization phase as a **cue-response mapping**. This refers to the stable mapping of how a learner initiates the next response path under specific situational cues. This formulation is used to locate and train repeatable bottleneck segments and does not deny the importance of meaning construction or representational processing.

From this perspective, training can be approximately described as two types of operations: establishing a viable next-response path for a new cue, or correcting/replacing an existing but inefficient or incorrect response path. This paper operationally defines "mastery" as follows: under the same or similar cues, the target response becomes faster, more stable, and involves a lower cognitive load. These trends are characterized by repeatable performance records [**?**, **?**, **?**].

Research Objectives: This paper proposes the **Micro-bottleneck Internalization Protocol (MBIP)** and details its components, constraints, and testable predictions. Its objectives include:

1. **Defining the i-Zone:** Focusing learning resources on the learner's current micro-breakpoints with the highest potential for gain, thereby reducing inefficient repetition of already mastered content.
2. **Constructing a Dual-Track Learning System:** The *micro-reinforcement track* is built around the i-Zone using MS, retrieval, explanation, and repetition. The *macro-organization track* maintains the retrievability of the overall system through macro-structural representations, such as structured outlines. The consistent mapping requirement dictates that the MS used for micro-reinforcement must have a clear corresponding position within the macro-structure. The dual-track approach emphasizes functional division rather than specific tools; in practice, it can be implemented via paper-and-pencil or digital tools. Digital tools are primarily used to reduce the costs of repetition and scheduling (e.g., short audio playback and optional spaced repetition scheduling).
3. **Proposing Standardized Internalization Operational Units:** Using a Minimal Scenario to trigger active retrieval, and employing Scene-Based Self-Explanation as a reusable "explanatory scaffold." Combined with spaced repetition, this enables low-cost repeated invocation. Furthermore, the protocol provides criteria for mastery, failure signals, and falsifiable predictions.

It should be noted that MBIP does not replace disciplinary instruction nor does it generate new knowledge. It focuses on process variables such as encoding, consolidation, retrieval, and organization. Its goal is to improve the efficiency per unit of time during the internalization phase and release cognitive resources for higher-level thinking activities.

Contributions of this Paper: Compared to traditional review methods, which struggle to consistently align training with individual bottlenecks in natural settings, this paper proposes **Scene-Based Self-Explanation (SBE)**. The content of the explanation is strictly centered on the next response path required by the Minimal Scenario (MS), and the output is compressed into a reusable, short explanatory scaffold. A typical medium for this is a learner' s self-recorded short audio: recording occurs simultaneously with card creation or explanation generation, and playback occurs during review at intervals to reduce the cost of reproduction. If environmental constraints make recording or playback unfeasible, it can degrade into a short text scaffold; this paper regards text as a viable but relatively sub-optimal implementation, with expected reductions in the advantages of initiation efficiency and error-correction speed. The audience for SBE is functionally defined as the "future self." Therefore, the explanation only needs to cover the minimum prerequisites necessary to initiate the next step after the trigger, without needing to provide full background for an external audience. This allows the explanation to remain stably focused on the "stuck step." For example, in a ten-step derivation or implementation, if the bottleneck occurs between steps seven and eight, the explanation may cover only the trigger cue for step seven and the key transition to step eight, without restating the remaining steps.

Regarding presentation, MBIP separates the "trigger cue (MS)" from the "explanatory scaffold (SBE),"enforcing a "retrieval first, correction second"sequence through procedural constraints. In practice, this can be implemented using a double-sided cue-scaffold medium: the cue side presents only the MS to trigger retrieval, while the scaffold side presents the SBE and necessary error correction or supplementary information for calibration. Accordingly, a failure signal can be operationalized as: the inability to initiate the next response within a limited time window, or the inability to proceed even after invoking the explanatory material. In such cases, a rewrite of the explanation is triggered, or the learner returns to the macro-structure to repair the chain. Theoretically, this design is compatible with existing mechanisms: first, SBE belongs to generative processing; second, oral production introduces a "vocalization/production" component; third, playing back explanations auditorily during the review phase provides learners with an input channel distinct from pure visual reading. Based on this, this paper does not presuppose that audio playback is inherently superior to text review, but rather proposes a testable prediction: when explanations are compressed into minimal scaffolds for the "future self" and the review task requires rapid initiation of the next response path, audio playback may be more advantageous for error correction and initiation efficiency per unit of time. This advantage is expected to diminish when the material relies primarily on elaborate symbolic visual information or when audio use is obstructed.

Furthermore, MBIP compresses review operations into the rapid invocation of a small number of MS units. By using a fixed "retrieval-then-correction" sequence, it significantly reduces initiation and interaction friction, making it easier to embed review sessions into stable, low-interference daily time windows. Under this framework, a pre-sleep review can be simplified into a brief micro-internalization cycle: for non-image-dependent MS, learners can complete retrieval via trigger cues (including optional short audio) and invoke the SBE on the answer side for correction when necessary. For image-dependent MS, the process is dominated by image retrieval, with audio primarily serving prompting and correction functions. Relevant boundary conditions and tool selections will be further discussed in Section 4.8.

Compared to review designs based on "knowledge points" or "problems," this paper shifts the minimal intervention unit of the internalization phase down to the **cue-response mapping**. By using MS to compress training triggers into the smallest directly invokable fragments, it reduces the redundancy and dilution caused by replaying entire problems. Compared to the macro-level ability intervals of the Zone of Proximal Development (ZPD), this paper operationalizes an individual' s current bottlenecks as a locatable **i-Zone** and provides observable identification cues to prioritize the allocation of training resources to breakpoints with higher gains per unit of time. Finally, rather than focusing solely on fragmented practice or structural organization, this paper proposes the principle of "micro-reinforcement track—macro-structure track" consistent mapping and failure backtracking rules. This constrains the long-term coupling between fragmented training and macro-structures, avoiding the separation where "fragments can be performed but the structure is not retrievable" or "the structure is complete but micro-steps cannot be invoked."

2.1 理论谱系与研究定位——从“如何学会”到“如何内化”

The field of the learning sciences has accumulated a rich and mature theoretical tradition to explain how humans acquire, understand, and apply knowledge. However, existing research exhibits a significant imbalance in focus: while a vast body of theoretical and empirical work centers on the "initial acquisition of knowledge," there is a relative lack of systematic operational frameworks addressing "how knowledge is stably internalized and maintained over the long term." Specifically, at the level of learner self-regulation, there is a dearth of executable frameworks capable of stably coupling multiple cognitive mechanisms into daily routines.

This paper focuses on the internalization stage, aiming to propose a standardized learning protocol that learners can repeatedly execute in their daily studies: the Micro-bottleneck Internalization Protocol (MBIP). The fundamental premise of this protocol is to isolate the microscopic "bottlenecks" —the specific points of friction within complex tasks that truly consume time and determine performance–and transform them into reproducible training objects. These objects are then organized into practice sessions through a chain of "minimal context–active retrieval–contextual self-explanation–spaced repetition." Furthermore, the protocol requires that this micro-level training maintains a consistent mapping with the macro-level knowledge structure [?]. To define the theoretical origins and necessity of this protocol, this chapter reviews the three major traditions of behaviorism, cognitive psychology, and social constructivism. By identifying the common gaps left by these traditions at the "operational layer" of the internalization stage, we provide the groundwork for the formal definition of the subsequent framework.

2.2 行为主义——强化逻辑与"任务颗粒度"问题

(1) Operant Conditioning and Programmed Instruction

Behaviorism emphasizes the role of reinforcement and feedback in shaping behavioral probability, leading to the development of practical principles such as programmed instruction (small steps, active response, immediate feedback, and individualized pacing) [?]. These ideas have exerted a profound influence on subsequent adaptive learning systems and computer-assisted instruction. In complex cognitive tasks, learners must engage in iterative cycles of active thinking and feedback adjustment. The key lies in identifying which specific cognitive links or steps require continuous practice and refinement.

(2) The Bottleneck of Behaviorism in Complex Cognitive Tasks: Unclear Reinforcement Units

In complex cognitive tasks–such as mathematical reasoning, physical modeling, or the construction of proofs–the primary challenge is often not whether "reinforcement is effective," but rather where the intervention unit of practice should be directed [?]. If the training unit is defined as the "entire problem" or the "final answer," it fails to correspond precisely to the learner' s actual cognitive breakpoints. Furthermore, it becomes difficult to distinguish between automated steps and critical decision-making nodes, leading to significant time wasted on low-gain repetition. This suggests that internalization training requires finer-grained, locatable intervention units. Correspondingly, MBIP employs the "Minimum Scene (MS)" as an operational training unit and utilizes the "i-zone" to continuously direct practice resources toward an individual' s most unstable micro-links. This makes the question of "where to direct the intervention" executable at the operational level.

(3) Media Conditions and Operational Precision Constraints: From "Historically Impossible" to "Currently Realizable"

Early media conditions limited fine-grained segmentation and dynamic scheduling. Even if smaller training units were theoretically necessary, it was difficult for learners to stably record, reproduce, and track changes in these units during natural learning. With the proliferation of digital tools, the preservation, rapid reproduction, spaced scheduling, and process recording of localized scenes have become more practically achievable.

It must be emphasized that the theoretical premises of MBIP do not depend on specific media or software. The crux lies in whether the Minimum Scene (MS) can be externalized and reproduced with low friction, whether explanatory scaffolds (such as scene self-explanation) can be presented on the answer side, and whether the system supports repeated invocation. In practice, as long as a medium can carry the necessary text and images and support audio recording and playback when needed, the basic operation of MBIP can be realized. This means that MBIP is not a method that "must rely on flashcard software"; learners can also use general-purpose tools such as PowerPoint to organize materials according to the Minimum Scene principle and achieve effective results.

Building on this, the differences between various media are primarily reflected in execution friction and scheduling capabilities. Paper-based notes or error workbooks can also be organized for training according to MS principles, but they typically involve higher friction regarding rapid reproduction, spaced scheduling, and historical tracking. Particularly when explanatory scaffolds are generated and replayed in audio form, purely paper-based media struggle to support these key reuse links and are generally not the preferred implementation. The marginal costs of reproduction and tracking are harder to maintain at a low level over the long term. In contrast, electronic media are often more conducive to high-frequency reproduction and precise tracking. For example, if learners wish to customize review sequences and presentation formats more freely, they can use tools like PowerPoint to organize MS; if they wish to automate spaced scheduling and due-date reminders through software, they can utilize flashcard applications.

Therefore, the choice of medium is not a theoretical component of MBIP, but rather an implementation condition that affects long-term adherence and operational precision. At the execution level, the problem transforms into whether the learner possesses self-executable rules to convert media capabilities into stable, daily training workflows. MBIP compresses these possibilities into protocolized steps, making fine-grained interventions easier to execute repeatedly and facilitating transitions between different platforms.

2.3 认知心理学——提取、负荷与生成机制

Cognitive Load and the Necessity of De-redundancy

Cognitive Load Theory posits that working memory capacity is limited; excessive information irrelevant to the goal increases extraneous load and impairs learning efficiency [?]. This suggests that the internalization phase should emphasize precise control over information structure. Accordingly, MBIP introduces the compression principles of Micro-Slices (MS) at the prompting stage, using minimal yet essential cues to trigger target responses. This reduces the occupation of cognitive resources by information unrelated to the current bottleneck. When compression leads to recognition failure, the system reverts to more complete materials and acquisition support to avoid conducting internalization training under incorrect entry conditions.

Retrieval Practice: Review Must Be Active

Extensive research supports the finding that retrieval practice is more conducive to long-term retention than repeated reading [?, ?, ?, ?]. However, practical challenges remain: how to select retrieval targets so they consistently align with unstable knowledge components rather than inefficiently reproducing mastered content. MBIP sets "retrieval first" as a process constraint. By utilizing i-zone localization and MS construction, it anchors retrieval targets to reproducible bottleneck scenarios, reducing the risk of practice being diluted by already proficient steps [?, ?]. In short, the MBIP strategy is not to increase the total volume of practice, but to improve the "hit rate" of each unit of practice on specific bottlenecks.

Generative Activities and Self-Explanation: The Gap Between Effect and Sustainability

Generative activities, such as self-explanation and learning-by-teaching, typically promote deeper processing, transfer, and long-term retention [?, ?]. However, these strategies often face sustainability hurdles in real-world learning: ill-defined explanation boundaries can lead to an expansion of output scope, resulting in high time costs and cognitive load, particularly in long-chain, high-order tasks. This occurs because the assumed audience for the explanation is usually another person; in such cases, the learner must provide shared prerequisite knowledge, further increasing the burden. Consequently, while existing evidence robustly supports *why* explanation is effective, there remains a lack of a standardized operational framework for *how* to implement explanation in a low-cost, reusable, and embeddable manner.

MBIP aims to fill this operational gap. It restricts the scope of explanation to local decision points corresponding to an MS and adopts "scenario self-explanation" as a fixed procedural step. MBIP functionally defines the audience of the explanation as the learner's future self.

This design allows for the radical compression of the explanation's scope without losing critical evidence, focusing solely on key reasoning at the micro-bottleneck. Naturally, this omission assumes that the learner can still reliably enter the scenario and that no systematic forgetting of prerequisite knowledge has occurred. This stands in stark contrast to explaining to others, where the lack of shared background necessitates a more comprehensive and costly presentation.

Building on this, MBIP focuses on concrete implementation. Audio recording is a typical, low-friction medium for realizing this "self-dialogue," though it is not the only one. To prevent explanation from degrading into a mere recitation of answers, MBIP introduces a proficiency criterion: the minimum standard is whether the next response can be independently generated under the same MS based solely on that explanation.

Furthermore, from an information-processing perspective, the review channel of explanatory materials may influence internalization efficiency. Compared to pure visual reading, auditory playback offers several potential advantages. First, vocal output is generally faster than typing during the generation phase, allowing for more semantic information to be expressed per unit of time; moreover, pauses, repetitions, and self-corrections during oral delivery often naturally expose unstable reasoning nodes. Second, during the review phase, replaying existing explanatory materials in auditory form can trigger path reconstruction without occupying visual resources, helping to maintain the continuity of retrieval practice in low-interaction contexts. Third, oral production and auditory playback form a "generate-listen" closed loop, making the explanatory material both a product of generative processing and a scaffold for subsequent retrieval.

It must be emphasized that this paper does not presuppose that auditory review is superior to text-based review for all tasks. When learning materials rely heavily on fine visual symbols (such as complex formulas, geometric structures, or image details), a purely auditory channel may be insufficient to support complete path reconstruction. Therefore, the advantages of auditory playback should be understood as a context-dependent implementation condition rather than a universal cross-task hypothesis.

From an operational standpoint, for the aforementioned mechanisms to function stably within daily self-scheduling, they are often constrained by executive friction. MBIP compresses review into the rapid invocation of a small number of MS and mandates a fixed "retrieval-then-correction" sequence. Consequently, with appropriate tool support, review can be embedded into low-interaction daily time windows (e.g., short-duration review scenarios on mobile devices). Relevant boundary conditions and tool selections will be further discussed in Section 4.8.

2.4 社会建构主义——从最近发展区到微观卡顿点

(1) The Zone of Proximal Development (ZPD) and Instructional Challenges

The Zone of Proximal Development (ZPD) emphasizes that learning should occur in the region where a learner “cannot yet complete a task independently but can succeed with scaffolding support” [?]. While this concept has profoundly influenced scaffolded instruction and collaborative learning, continuously and accurately locating an individual’ s ZPD in self-regulated learning or large-scale scenarios remains highly dependent on teacher experience, making it difficult to standardize or scale. Regarding the internalization stage, the critical difficulty lies in the fact that after achieving initial understanding, learners must independently maintain the focus of their practice during daily repetitions. In reality, however, external experts are not available to provide continuous, step-by-step immediate scaffolding and diagnosis for every repetition. The response of the Micro-Scaffolding Based Internalization Protocol (MBIP) is to transform the “scaffolding” component into self-executable procedural constraints for the learner—such as prioritizing retrieval, generating or invoking contextualized self-explanations at points of failure, and returning to more comprehensive materials when entry conditions are not met—while maintaining a channel to return to external scaffolds (teachers, peers, AI, or examples) if entry conditions remain unsatisfied.

(2) From Descriptive Concepts to Operational Signals

As a descriptive theoretical framework, the core challenge of the ZPD lies in its operationalization. While it defines an optimal “zone” for learning, it rarely addresses the micro-level segments where a learner specifically becomes “stuck” during a task. Training during the internalization stage urgently requires micro-targets that can be self-identified and dynamically tracked by the learner.

MBIP addresses this by converging expressions and reducing theoretical commitments: this paper does not advocate for the precise measurement or alternative definition of the ZPD, but instead proposes an “approximate positioning method” better suited for self-scheduling. Specifically, learners continuously produce observable performance signals during task execution, such as the repeated occurrence of the same type of error, significantly prolonged response times at a critical step, or obvious stagnation at a particular stage requiring repeated consultation of materials. While these phenomena are not strictly equivalent to formal ZPD indicators, they serve in practice as low-cost cues that “scaffolding is still required here” or “stable internalization has not yet been achieved.” To facilitate repeated use within the protocol, MBIP summarizes this set of signals as the “i-Zone” : the micro-segments where a learner repeatedly exhibits instability, error-proneness, or high cognitive costs within a specific task context. The function of the i-Zone is to transform the question of “where scaffolding is needed” into directly perceivable task performance, thereby providing sustainable direction for internalization training.

Based on this, MBIP transforms the problem of locating the ZPD into the problem of identifying the i-Zone. It operationalizes the "current minimal point of stagnation" through the concept of the i-Zone and then solidifies it into a reproducible training object via Minimal Scenarios (MS). Simultaneously, by requiring that each MS has a clear position within the macro-knowledge structure (such as mind maps or outlines), the protocol ensures that these micro-level interventions serve the construction and retrieval of the overall knowledge framework.

2.5 小结——理论整合的空白与研究问题提出

A synthesis of related work across the three traditional domains reveals that the internalization stage still suffers from three recurring integration gaps that have yet to be unified or resolved:

1. **The Unit of Analysis Gap:** There is a lack of a clearly defined minimum unit of intervention within complex tasks.
2. **The Localization Gap:** There is a lack of methods to consistently focus resources on the specific micro-processes where an individual has the greatest potential for gain.
3. **The Integration Gap:** There is a lack of a protocol that integrates reinforcement, retrieval, generation, and structural construction into a format that learners can execute independently.

Furthermore, constraints regarding the cost and reusability of self-explanation make it difficult for such integration to be stably implemented [?]. It is important to emphasize that these gaps do not deny the existence of the underlying mechanisms; rather, they point to the absence of a combinatory protocol that is both self-executable by the learner and capable of consistently aligning with micro-level bottlenecks.

Based on these observations, this paper proposes and systematically constructs the Micro-Bottleneck Internalization Protocol (MBIP). Chapter 3 will provide formal definitions of the core constructs, while Chapter 4 will outline the standard operating procedures and demonstrate the methods of implementation.

2.6 命题与证据传统的对齐口径（证据地图）

To demonstrate that P1–P3 are not merely empirical aggregations, this paper provides an "evidence alignment framework" designed to compare established research traditions and subsequent testable questions within a unified structure. This framework serves to standardize evidence traditions and metric specifications, facilitating the formulation of testable research questions for future study; notably, this paper does not report quantitative results from completed reviews.

(1) Search Criteria and Inclusion Standards

To facilitate subsequent scoping searches and systematic coding, this section outlines a simplified version of the primary search and inclusion criteria. The comprehensive search strings, along with detailed screening and coding protocols, will be documented and provided in the supplementary materials.

Databases: The search plan covers core databases in the fields of educational psychology, cognitive science, and the learning sciences, including PsycINFO, ERIC, and the Web of Science Core Collection.

Organization of Search Terms: The search terms will be organized around three categories of mechanistic evidence: prompting and information de-redundancy (corresponding to P1), the timing of extraction and feedback (corresponding to P2), and self-explanation and generative output along with its sustainability (corresponding to P3). Detailed search term lists and their combinations are provided in the Supplementary Materials.

Inclusion and Exclusion Criteria: The literature to be included primarily consists of studies containing empirical data (such as experiments, quasi-experiments, and classroom intervention studies) as well as high-quality systematic reviews. Purely opinion-based articles, as well as studies that only report learner attitudes or satisfaction without measuring learning outcomes, will be excluded.

Evidence Analysis Framework: Subsequent evidence synthesis will be conducted by screening and coding according to the dimensions shown in . This approach ensures that different evidence traditions can be aligned and compared within a standardized framework.

Table 1: MBIP Evidence Analysis Framework (Research Agenda Type; used to anchor evidence traditions and standards, not for reporting quantitative results)

Identifiable Operational Characteristics in Literature	Corresponding Proposition	Prompts/Material Segmentation (Training Units)
Segmentation of Training Units	Breaking down a complete task into smaller, trainable units; reducing information irrelevant to the current step; emphasizing that "cues are just sufficient to initiate the next step."	P1

Identifiable Operational Characteristics in Literature	Corresponding Proposition	Prompts/Material Segmentation (Training Units)
Practice Sequence (Do-then-See)	Requiring the learner to attempt an answer or generate a step first before providing the answer or explanation; limiting opportunities to "view the explanation first."	P2
Explanation and Error Correction (Local Focus)	Requiring learners to explain "why this local step is performed this way" or "how the next step is initiated"; the scope of explanation can be restricted to a local scenario; the medium can be oral or short text.	P3

结果口径（怎么判断更好）报告延迟保持、迁移、无提

P1–P3 demonstrate performance in independent tasks; alternatively, they report metrics related to the learning process, such as error types and completion times.

3.1 引言

Following the analysis of the "operational-level gap" in the knowledge internalization stage presented in the previous chapter, this chapter provides a formal definition of the core theoretical components of the Micro-bottleneck Internalization Protocol (MBIP). Departing from empirical learning suggestions or general tool-usage guidelines, this chapter aims for operationality, testability, and reproducibility by defining the key concepts, units of analysis, and functional mechanisms within the system.

The theoretical framework of MBIP consists of three core propositions that are mutually constraining and supportive. Together, they form a self-executable combinatorial protocol for learners, designed to align practice resources more stably with individual micro-bottlenecks during the internalization phase, while maintaining a consistent mapping between micro-level training and macro-level knowledge structures. The following sections will elaborate on these three propositions and provide their operational definitions within the MBIP framework.

3.2 命题一（P1）：微观瓶颈原则（The Principle of the Micro-Bottleneck）

(1) Proposition Definition Proposition 1 (P1): Internalization training is more likely to benefit from decomposing training units down to the “minimum bottleneck” segments where learners repeatedly encounter friction in specific tasks, rather than defaulting to complete problems or entire passages as the primary training units.

This proposition addresses the “unit of analysis gap” identified in Chapter 2. Its purpose is not to deny the importance of holistic understanding or representation construction, but rather to provide an operational object for the internalization phase that is easier to locate, reproduce, and track. (2) Mechanism Elaboration: The Minimum Scene (MS). To operationalize P1, the MBIP introduces the core construct of the “Minimum Scene” (MS).

Definition 3.1: Minimum Scene A Minimum Scene is defined as the smallest set of cues within an acceptable time window that is sufficient for a learner to stably identify the current context and trigger the subsequent response path.

This set of cues must satisfy two constraints: 1. Necessity: If the cues are fewer than those in this set, the learner will be unable to stably identify key contextual cues, making it difficult to initiate the target response path. 2. Non-redundant Minimum Sufficiency: If the cues exceed this set, they are more likely to introduce information irrelevant to the target segment, thereby increasing extraneous cognitive load and weakening diagnostic precision.

Therefore, the goal of the MS is not to reconstruct the “complete problem,” but to retain only the contextual cues necessary to trigger the bottleneck, making the point of friction easier to repeatedly retrieve, interpret, and practice.

Entry Conditions (Minimum Caliber): The learner must at least be able to state “what kind of scene this is” or “what type of operation the next step belongs to.” If these entry conditions are not met, the learner should return to more comprehensive materials or external scaffolding (such as examples or feedback from teachers, peers, or AI) to correct foundational understanding before entering MS training. (3) Function and Implementation in MBIP. Functional Positioning: The MS serves as the atomic unit for retrieval, interpretation, feedback, and tracking within the MBIP. It is designed to concentrate practice resources on a single bottleneck segment (the i-Zone) while aiming to reduce extraneous load unrelated to that specific bottleneck.

Implementation Caliber: The essence of an MS is its information structure rather than a fixed medium. A common implementation is to format it as a prompt card to facilitate the rapid reproduction of the same localized context. The minimum operational standard for determining its validity is whether the learner can initiate the target response path and proceed to the next step based solely on the information provided.

Brief Comparison Between Minimum Scenes and Traditional Learning Units (Design Caliber) Traditional units typically contain multiple steps and sequences, whereas the MS focuses as much as possible on a single bottleneck segment. While traditional units provide complete information—often including redundancy and interference—the MS ensures cues are "just enough" to facilitate rapid reproduction.

3.3 命题二（P2）：主动提取原则（The Principle of Proactive Retrieval）

(1) Proposition Definition Proposition 2 (P2): In exercises targeting any bottleneck stage, the workflow requires an initial attempt at active retrieval prior to accessing explanations or feedback. This minimizes the substitution of retrieval by "viewing explanations first" [?].

This proposition addresses the "gap localization" issue identified in Chapter 2: it utilizes repeatable signals of retrieval failure to continuously align practice resources with unstable cognitive links. (2) Mechanism Elaboration: Attempt First, Correct Later. A substantial body of research demonstrates that retrieval practice is generally more beneficial for delayed retention and unaided recall than passive review methods such as re-reading [?][?][?]. P2 operationalizes this as a procedural constraint: retrieval must be attempted before exposure to explanations or feedback.

In the absence of a prior attempt, viewing answers or explanations is more likely to primarily enhance feelings of familiarity and recognition cues; however, the ability to recall information in unprompted contexts still requires training through retrieval attempts. Even when retrieval fails, the initial attempt serves two diagnostic functions: it explicitly exposes the learner' s sticking points and error types, providing a targeted focus for subsequent explanation and error correction. Furthermore, it serves a priming function: retrieval failure creates a distinct "cognitive gap," making the subsequent explanation more likely to be used to bridge critical links rather than being passively browsed. (3) Operationalization in MBIP. In the Model-Based Instruction Protocol (MBIP), P2 is implemented through the functional separation of prompts and explanations: the prompt area is used exclusively to trigger retrieval, while the explanation and feedback area is presented only after the retrieval attempt to provide explanations, feedback, and necessary corrective cues. Scenario-Based Explanation (SBE) is the core component of the explanation and feedback area; however, it is not equivalent to a "standard answer." Its function is to help the learner restore and stabilize the activation of the "next-step response path."

At the implementation level, this separation can be achieved through various media. For instance, in a flashcard system, a "front-back" structure can be employed: the front presents a Minimal Scenario (MS) to trigger retrieval, while the back provides the Scenario-Based Explanation and necessary supplementary information. In presentation software (such as PPT), this can be achieved through adjacent slides or step-by-step animations to control the workflow of "viewing the problem first, then the explanation." Differences between various tools primarily manifest in presentation styles and operational costs, without altering the underlying logical structure of the protocol.

3.4 命题三 (P3)：场景自我解释原则 (The Principle of Scene-Based

Self-Explanation) (1) Proposition Definition Proposition 3 (P3): At the bottleneck stage where retrieval fails, the learner must generate a self-explanation focused on that Minimal Scenario (MS) and utilize it as a core material for subsequent reproduction.

This proposition addresses the "integration gap" identified in Chapter 2: it attempts to transform generative activities from high-cost, sporadic strategies into low-cost, reusable, and sustainable internalization processes. (2) Mechanism Elaboration: Scene-Based Self-Explanation (SBE). To operationalize P3, MBIP introduces "Scene-Based Self-Explanation," which refines traditional self-explanation in two key dimensions: Narrowing the object of explanation: The explanation is strictly limited to the local decision point corresponding to the MS, answering "In this state, how is the next step initiated, or what is the key basis for it?" Setting the target audience: The audience for the explanation is functionally set as the learner' s future self. This allows the explanation to omit irrelevant background information without losing key evidence, thereby reducing output costs and improving sustainability. (3) Implementation and Reuse Logic. SBE must be solidified into explicit material that can be repeatedly invoked to stably trigger the same "cue-response path mapping" during subsequent reviews. At the implementation level, if environmental conditions permit, MBIP prioritizes generating and saving SBE in an audio format.

The rationale for using an audio format is primarily based on considerations of execution efficiency and processing mechanisms. First, from the perspective of execution cost, oral production generally has a higher rate of information expression, making it more conducive to completing explanation generation within a given unit of time compared to text input. More importantly, the pauses, hesitations, and self-corrections that occur during oral narration often serve as real-time performance signals of unstable links, helping to expose micro-bottleneck locations.

Second, from the perspective of cognitive processing, oral explanation itself constitutes an instance of active retrieval and generative processing. Reconstructing local decision logic in the learner's own language is more likely to strengthen the stable mapping between specific situational cues and the subsequent response path than mechanical repetition of the original text. This, in turn, enhances the feasibility and degree of automation in future retrieval.

Furthermore, during the review phase, audio materials can be played back directly through the auditory channel. After attempting "retrieval first," the learner only needs to invoke the existing explanatory material to quickly restore the previously generated path structure, thereby achieving a low-friction cycle of "generate once, reuse many times." In contexts where the visual channel is already occupied with symbol recognition or diagrammatic processing, auditory playback may also alleviate the information processing load to some extent.

It should be noted that this paper does not regard the audio format as a theoretical necessity. In situations where overt vocalization is inconvenient or where tasks are highly dependent on precise visual symbols, brief text can serve as an alternative implementation. However, compared to audio, textual review typically involves re-reading and re-organization, which may introduce additional processing costs. The relative advantages and boundary conditions of different media remain subjects for further empirical testing.

Regardless of the medium used, the review process must follow the sequential constraint of "retrieval first, then invocation of explanation." This prevents the explanatory material from operationally replacing active recall, which would otherwise weaken the diagnostic function of the internalization process. (3) Function and Implementation in MBIP. SBE exists as the core content of the explanation and feedback zone. It is not equivalent to a "standard answer" but is rather a local scaffold created by the learner for their future self. Its goal is to help the learner restore and initiate a stable mapping of the next response path under the same situational cues. Its minimum standard can be expressed as follows: can the learner independently generate the next response in a similar MS relying solely on this scaffold (the SBE and its necessary minimal auxiliary materials)? It should be noted that while the SBE is the core of the back-side scaffold, minimal necessary information (such as key diagrams or symbolic conventions) may be supplemented in certain tasks to reduce reproduction friction and improve reusability.

3.5 三大命题的系统协同与流程模型

P1, P2, and P3 are not isolated principles but rather form a coupled system of mutual prerequisites. P1 defines the training object: the "Minimum Stuttering Point" (MS) serves as the fundamental unit of training.

P2 defines the sequence of practice: retrieval must be attempted before exposure to explanations or feedback. P3 defines the handling of failure: at the point where retrieval fails, a Self-Back-Explanation (SBE) is generated or invoked to provide reusable explanations and error-correction pathways.

3.6 本章小结

This chapter provides a formal definition of the theoretical core of the Micro-Bottleneck Internalization Process (MBIP). It proposes and elaborates on three core propositions: the Micro-Bottleneck (P1), Proactive Extraction (P2), and Scenario Self-Explanation (P3). Together, these three elements constitute the internal logic of MBIP.

This framework utilizes reproducible local bottlenecks as fundamental training units. It employs a "prior extraction" mechanism to reveal whether the system can independently generate the subsequent step. When extraction fails, the system generates reusable explanations and error-correction cues in a low-friction manner, while maintaining a consistent mapping between micro-level training and macro-level structures. Building upon this foundation, the next chapter will present the Standard Operating Procedure (SOP) and typical implementation examples.

4.1 引言：协议的性质与操作目标

This chapter translates the core propositions proposed in Chapter 3 into a set of testable protocol hypotheses oriented toward the internalization stage. These include operational rules, minimum functional criteria, and failure triage logic. These hypotheses are intended to guide the subsequent organization of the evidence map and research design, rather than to assert universally proven conclusions.

The operational objective of MBIP is as follows: after the learner completes initial acquisition, practice resources are converged from broad knowledge points or full-problem reproduction into smaller, reproducible training units. Specifically, the goal is to determine whether the next response path can be initiated within an acceptable time window given specific situational cues. This chapter organizes the protocol according to three categories of constraints: the identification and encapsulation of the minimum training object; the process by which generative explanations are solidified into reusable materials; and the consistent mapping and failure backtracking rules between micro-practice and macro-structures. An appendix is provided to offer a complete execution example, demonstrating how the protocol is implemented within a specific task.

The main text focuses on the constituent elements of the protocol, the minimum functional criteria, and the failure triage logic, without expanding into multi-case comparisons or detailed tool specifications.

4.2 模块一：最小训练对象的界定与封装（Unit Definition and Encapsulation

This module establishes three conceptual divisions and their corresponding minimum functional criteria. These are used to define the training objects and generate observable loss functions.

1. Unit of Analysis: Cue-Response Path Mapping

The fundamental unit of analysis is defined as the mapping of "cue-response" paths. In the context of machine learning and behavioral modeling, this involves identifying the specific environmental or data-driven stimuli (cues) and the subsequent computational or behavioral outputs (responses). By isolating these paths, we can precisely delineate the boundaries of the training process and establish a rigorous framework for evaluating model performance.

2. Functional Criteria and Training Objects

To ensure the effectiveness of the encapsulation, we define minimum functional criteria that each training object must satisfy. These criteria serve as the baseline for determining whether a model has successfully internalized the required mappings. By setting these thresholds, we can systematically categorize training objects based on their complexity and the specific requirements of the task at hand.

3. Generation of Observable Loss

The final component of this module involves the generation of observable loss. By comparing the predicted response of the cue-response path against the ground truth or desired outcome, we derive a quantifiable metric of error. This loss function is essential for the iterative optimization of the model, providing a clear signal for the adjustment of parameters during the deep learning process. The observability of this loss ensures that the training progress can be monitored and validated against the established functional criteria.

分析单位被表述为：在特定线索出现时，学习者能否在可接受的时间窗口内启动并

The next step in executing a goal (e.g., a specific step in a derivation, a transformation, or a debugging operation). (2) Priority Target: The i-Zone as a Bottleneck Signal. The i-Zone refers to recurring signals of stagnation during task execution, typically characterized by repetitive errors, significantly prolonged reaction times, or cognitive paralysis where progress ceases. The function of the i-Zone is to help learners continuously direct their practice resources toward the "micro-steps most in need of repair."

Note: Although the intuition behind the i-Zone is inspired by the boundary focus of the Zone of Proximal Development (ZPD), in this paper, it is used solely as an empirical clue or approximate indicator for locating bottlenecks in self-directed learning scenarios. It should not be equated with the formal definition of ZPD within the context of instructional scaffolding. (3) Explicit Carrier: The Minimal Scenario (MS) serves as the explicit carrier that "packages" a bottleneck into a reproducible prompt.

Operational Rule 4.1 (MS Encapsulation Constraints): The construction of an MS aims for the goal of "just enough cues" : - Insufficient cues: Leads to an inability to recognize the scenario, preventing the initiation of the next step. - Excessive cues: Increases the likelihood of introducing information irrelevant to the current step, thereby consuming attentional resources and diluting diagnostic precision.

Minimum Functional Criterion: Based solely on this prompt, can the learner enter the same type of local context and begin attempting the next step? Failure Branching 4.1 (Entry Conditions Not Met): When a learner cannot even initiate the process—failing to identify "what kind of scenario this is" or "which category of operation the next step belongs to" —entering internalization training under these conditions is generally inappropriate. In such cases, the learner should first return to more comprehensive materials or external scaffolding (such as examples, peer feedback, or teacher guidance) to satisfy the entry conditions. After comprehension has been corrected, the learner may then return to the MS for internalization.

4.3 模块二：生成性解释的受众重定向与低摩擦固化（Re-orientation

While generative explanation (Self-Explanation and Low-friction Fixation) is theoretically supported, it often faces two operational obstacles in real-world learning: first, the excessive cost caused by the expansion of the explanation' s scope; and second, the difficulty of reusing the generated explanations. MBIP addresses these issues through two key operational convergences:

Operational Rule 4.2 (Future-Self Orientation): The audience for the explanation is functionally defined as "one' s future self." This setting assumes that the recipient of the explanation shares the same learning path and error history as the current learner. Consequently, the explanation can focus specifically on "why this was done here" or "how to initiate the next step," allowing for the omission of extensive background information without losing critical evidence:

What are the key discriminative cues? What are the common causes of error? What retrieval cues should be used to extract the procedural path when encountering similar scenarios? Operational Rule 4.3 (Explanation as Material + Mastery Criterion): The resulting explanation is directly solidified as the review material for that Micro-Step (MS), such as a short audio clip or a brief text.

Minimum Mastery Criterion: Can the learner independently generate the next response for a similar MS by only viewing or listening to the explanation? If this criterion is not met, the learner must return to "Entry Condition Repair" or supplement more upstream prerequisite relationships.

Carrier Caliber (Low-friction rather than Fixed Medium): This paper defines the requirements for the carrier as "low-friction," meaning it has low startup costs and is easy to reuse. Audio recording is a common implementation method; in contexts where voice recording is inconvenient, short text can serve as an alternative. It should be noted that these alternative carriers are not presumed to be fully equivalent; the relative advantages and boundaries of different media are left for future research to examine.

4.4 模块三：微观—宏观的双轨耦合规则 (Dual-track Coupling Rules)

MBIP adopts a dual-track functional structure rather than a single-dimensional practice approach. The Micro-track is built around the i-Zone to construct Memory Slots (MS) and execute retrieval-feedback loops, with the objective of stabilizing local response pathways.

The Macro-track utilizes outlines or concept maps to maintain the availability of prerequisite relationships and the overall structural integrity. It should be noted that the Macro-track itself is not the original contribution of this paper; rather, this study focuses on achieving consistent mapping and low-friction alignment between micro-training units and the macro-structure.

Operational Rule 4.4 (Consistent Mapping: Index Alignment): Each MS must occupy a clearly defined position (such as a node or entry number) within the macro-structure.

Implementation Method: A low-friction implementation involves back-writing (copying) the target response of the MS in its original form into the corresponding node of the macro-structure. This reduces structural alignment costs and mitigates the risk of version drift. The specific tasks and learning stages where the benefits of this method are most pronounced remain to be defined by future research.

Operational Rule 4.5 (Failure Backtracking: Structural Shunting): During spaced repetition, if the "retrieval followed by explanation invocation" sequence fails to progress, it should be prioritized as a signal that "prerequisite relationships are unavailable." In such cases, current micro-practice must be suspended to initiate a macro-structure backtrack: locate the missing upstream concepts or rules, repair them, and then return to the MS.

4.5 实施条件与工具地位 (Implementation Conditions and

Instrumentality) A critical implementation requirement for MBIP is a "low-friction external carrier." Its fundamental functions are to present the MS, store and retrieve explanatory materials (such as audio or text), and support repetitive reproduction. Whether features such as automated spaced-repetition scheduling or process logging are utilized is considered an extension at the implementation level rather than a mandatory component of the protocol.

The tools can take various forms, such as flashcard systems, note-taking systems, or presentation software (e.g., PPT) capable of displaying images and playing audio. The logic of the protocol does not depend on specific software; the protocol focuses on workflows and constraints rather than specific software interfaces or operational steps.

Environmental Adaptability: The recording generation phase has specific environmental requirements (vocalization is necessary), whereas the playback phase has lower environmental requirements (e.g., using headphones). Consequently, the choice of carrier can be adapted or substituted based on the specific context.

4.6 关键过程信号的操作性界定（Operational Definitions）

To facilitate subsequent research and practical documentation, this paper does not describe the operational status of the Micro-Step Intervention Protocol (MBIP) using abstract terminology. Instead, it compresses these states into several directly observable "process signals." These signals are used to describe whether the protocol chain has been triggered and at which stage the learner encounters an obstruction; they do not, in themselves, constitute verified outcome conclusions.

(1) Response Latency Definition: The duration from the presentation of a Micro-Step (MS) cue until the learner outputs the first "effective initial action." An effective action refers to behavior that advances the subsequent response path (such as beginning to write a critical step, starting to verbalize the next operation, or entering key code), rather than merely flipping through materials or repeatedly reading the problem.

Purpose: Used to determine whether "the MS can stably trigger the target path." If significant pauses occur across multiple replications, it typically suggests that the path has not yet been internalized or that the MS cue is not yet "just sufficient."

(2) Failed Retrieval Signal Definition: The situation where the next step cannot be initiated within a limited trial period, or where initiation occurs but immediately leads to an obvious error that cannot be self-corrected.

Purpose: Used to identify the i-Zone. If the same type of failure occurs repeatedly under similar MS cues, it indicates a stable bottleneck. Conversely, if the location of the failure shifts frequently, it is more likely that the entry conditions are unstable or that there is a disconnection in the macro-structure.

(3) Explanation Dependence Definition: Whether the learner must invoke a situational self-explanation (verbal or textual) or other external prompts to advance when completing the target next-step response.

Purpose: Used to distinguish between "knowing but unable to retrieve" and "not yet having formed a path." When explanation dependence remains high over the long term, it often means the path has not yet been automated; a decrease in dependence usually signifies that the response path corresponding to that MS is being internalized.

(4) Structural Break / Index Loss Definition: When reproducing a specific MS, the learner is unable to identify its position within the macro-structure (e.g., which chapter node or conceptual branch it belongs to, or which upstream rules it depends on) or is unable to trace back to its prerequisite relationships.

Purpose: Used to trigger the "structural shunting" described in Section 4.5. This involves prioritizing the suspension of micro-exercises to return to the macro-structure for link repair before returning to the MS.

4.7 研究议程：可检验主张与边界条件（Research Agenda）

This section does not present the Micro-Step Based Iterative Practice (MBIP) as a proven, definitive conclusion regarding effectiveness. Instead, it is framed as a set of empirical propositions that can be tested and potentially falsified. The core research question is not simply "whether a tool is effective," but rather: when training is organized according to the three propositions of MBIP (P1–P3) and the dual-track consistent mapping rules, do the learner's bottleneck localization and path initiation exhibit a more stable and traceable trajectory of improvement?

(1) Testable Claims Regarding MS Encapsulation (Corresponding to P1)

If the "just enough cues" encapsulation of a Micro-Step (MS) holds true, then for the same investment of practice time, a learner's practice is more likely to hit bottleneck segments directly rather than being diluted by steps they have already mastered. Observable performance at the operational level includes: - More stable initiation under similar MS conditions (reduced initiation pauses). - Failure signals concentrated on a few specific steps rather than being scattered throughout the entire problem. - Explanatory materials that remain concise and precisely aligned with "how to initiate the next step."

Counter-examples and Boundaries: If entry conditions are not met (e.g., the learner cannot recognize the scenario or does not know "which category of operation the next step belongs to"), MS training may be ineffective or produce misleading failure signals. In such cases, the learner should return to more comprehensive materials or external scaffolding to correct their understanding.

(2) Testable Claims Regarding the "Retrieval-First" Process (Corresponding to P2)

If the "attempt first, correct later" process constraint is consistently executed, failures are more likely to manifest as repeatable diagnostic signals, making the *i*-Zone easier to identify and track continuously. Observable performance at the operational level includes: - Stalling points that recur consistently (rather than getting stuck in different places each time). - Stable error types, facilitating the generation of targeted scenario self-explanations. - Explanatory content focused on trigger cues and key transitions rather than expanding into full-length paraphrasing.

Boundaries: If learners habitually "look at the explanation first" during reproduction, failure signals will be masked. This leads to a distortion in *i*-Zone localization, making it difficult to keep training resources aligned with actual bottlenecks.

(3) Testable Claims Regarding Dual-Track Consistent Mapping (Micro-Macro Coupling)

If every MS maintains a stable index within the macro-structure and triggers structural backtracking upon failure, micro-level training is less likely to become fragmented and more likely to support path recovery during comprehensive tasks. Observable performance at the operational level includes: - A reduced sense of "disconnection" during practice (e.g., being able to complete individual cards but unable to advance the overall problem). - When an MS cannot proceed, the learner can more easily locate and repair missing upstream concepts. - A decrease in the frequency of structural positioning failures over time.

Boundaries: If the macro-structure is not maintained over the long term or if the cost of index alignment is too high, micro-cards are prone to drifting into a collection of fragments. This results in the "fragmentation trap," where a learner can recall or understand individual components but cannot retrieve the necessary path during real-world tasks.

4.8 睡前移动端的低交互复习：适用边界与工具选择

In practice, some learners choose to conduct review sessions on mobile devices before sleep to reduce the activation cost and improve daily compliance. For non-image-dependent Memory Structures (MS), the review process can be highly "auditory-based": learners first listen to the front-side trigger audio to complete mental retrieval, followed by the back-side explanatory material for verification. This approach significantly reduces visual strain and interaction costs. Conversely, for image-dependent MS (such as geometric figures, diagram structures, or symbolic details), viewing the image remains necessary to complete retrieval, with audio primarily serving as a prompt or for correction. While using a Bluetooth controller for page-turning—whether through a flashcard system or a slide presentation—can further lower interaction costs and enhance continuity, the benefits of this method are similarly constrained by the task's inherent reliance on visual information.

This practice demonstrates that MBIP does not depend on specific software; rather, its operation is highly dependent on whether the medium can present the MS and invoke explanatory materials with low friction. The differences between various tools lie primarily in their execution costs rather than in the underlying logic of the protocol itself.

5.1 局限性

(1) Environmental and Carrier Constraints: Low Friction is Not Universal (Environmental Requirements during the Generation Phase). Several key

stages of Micro-Bottleneck Intervention Protocol (MBIP) rely on low-friction external carriers to support high-frequency repetition and material reuse. Voice recording serves as a typical implementation; however, its generation phase requires overt speech, which may be difficult to execute in quiet environments such as libraries or study rooms. This limitation reduces execution frequency and long-term sustainability, implying that "low friction" is context-dependent:

Carrier selection must be viewed as a prerequisite for implementation rather than a default equivalence [?]. (2) Attribution Threshold: Zone i Localization Depends on Self-Diagnostic Ability. MBIP utilizes Zone i as a bottleneck signal to guide resource allocation. Slicing this zone requires learners to further operationalize "errors, slowness, or stagnation" into trainable micro-segments (such as discriminant conditions, rule invocation, or step initiation points). If the slicing is inaccurate, Micro-Scenarios (MS) and explanatory materials may accumulate around false bottlenecks, making it difficult to translate investment into performance gains. The introductory stage typically requires demonstration and periodic external calibration to reduce the risk of misjudgment [?, ?]. (3) Scope of Application: Task Structure Dependency and Decomposability Boundaries. The micro-training units of MBIP are cue-response path mappings under specific cues, using MS as reproducible trigger units. Consequently, it is better suited for learning tasks with relatively clear steps, diagnosable errors, and operationalizable subsequent response paths. For creative tasks such as writing, where evaluation dimensions are multifaceted and response paths are difficult to encapsulate stably, the benefits and operability may decrease. For skills heavily dependent on somatosensory and immediate perceptual feedback—such as athletic movements, musical performance, or painting—key bottlenecks often lie in motor control and calibration. While the concept of "micro-bottleneck localization" can be borrowed, the carrier and evaluation system must be reconstructed rather than directly applying the verbal Q&A flashcard paradigm [?]. (4) Insufficient Evidence Base: Currently Based Primarily on Protocolized Hypotheses. To date, support for MBIP stems more from theoretical integration and prototyped practical experience, lacking rigorous controlled trials and systematic evidence across tasks and populations. Therefore, this paper is better regarded as a "normative operational framework + falsifiable predictions" rather than a methodological conclusion with confirmed efficacy. (5) Risk of Solidifying Erroneous Explanations: Explanatory Materials Require External Validation. MBIP emphasizes high-frequency repetition and the reuse of explanations. If the explanatory material itself contains incorrect rules or faulty discriminant cues, repeated invocation may reinforce erroneous cue-response mappings, increasing the cost of subsequent correction. Thus, the explanation generation phase requires basic external validation (e.g., textbook examples, standard solutions, peer/teacher feedback, or AI feedback). (6) Institutional and Situational Constraints: Implementation Thresholds in High School Environments. The low-friction implementation of MBIP typically relies on overt vocalization of explanations and electronic devices to support high-frequency repetition. In certain high school study halls and classroom settings, silence regulations and device management raise the threshold for execution, thereby reducing daily compliance and the replicability and coverage within conventional school contexts [?]. (7) Structural Fit Limitations: Limited Coverage of Minimum Scenario Segmentation. Even in subjects with relatively clear step structures like mathematics or programming, practical observation shows that not 100% of problems or knowledge points can be stably segmented into reproducible Micro-Scenarios (MS). In the author' s practical sample, a small portion of items (approximately 5%) were difficult to encapsulate stably according to MS principles.

Therefore, this paper does not assume that all problems can be conveniently segmented into MS. When MS segmentation is infeasible or excessively costly, a more appropriate approach is to use full-problem review, complete derivation playback, or other revision methods as supplements, rather than mechanically applying the protocol process. (Note: The aforementioned proportion is derived from practical observations of specific subjects and samples; it serves only to illustrate the boundary of "non-universal coverage" and does not constitute a universal statistical estimate.)

5.2 未来研究展望

Future Research Directions

1. Conducting Rigorous Empirical Testing

Future research should employ controlled experiments or authentic classroom settings to test the falsifiable predictions of the MBIP framework under conditions of equal time, equal practice attempts, and equivalent feedback. Studies should simultaneously report outcome measures—such as delayed retention, unprompted retrieval, transfer performance, and reaction time—alongside process indicators, including failed retrieval rates, hint dependency, and structural fragmentation. Such comprehensive data collection is essential to verify whether the proposed protocol chain is effectively triggered and to identify its specific boundary conditions.

2. Developing "Structure-Card Integrated" Low-Friction External Carriers

To reduce the maintenance costs associated with dual-track alignment and material reuse, future work should explore the integration of the macro-structural track (e.g., structured outlines or concept maps) and the micro-reinforcement track (MS units and explanatory materials) within a single external system. This integrated carrier should allow each MS unit to be stably anchored to its corresponding structural node while supporting low-friction access to explanatory materials and automated spaced-repetition scheduling. It is important to emphasize that such a system serves as an implementation facilitator rather than a core theoretical component. Its primary value lies in lowering execution costs, preventing structural drift, and providing a robust foundation for the collection of granular process data.

5.3 本章小结

This chapter summarizes the primary limitations of Model-Based Inquiry Protocols (MBIP) across several dimensions, including the execution environment, attribution thresholds, task boundaries, phase aliasing, combinatorial attribution, and insufficiency of evidence. Based on these challenges, future research should prioritize two strategic directions. First, empirical testing must be conducted under strictly controlled conditions to generate cumulative evidence that links results with processes. Second, researchers should explore low-friction external carriers that integrate structure and scaffolding (cards) to reduce protocol execution and long-term maintenance costs. These advancements are essential to support the sustainable implementation of MBIP within authentic learning contexts.

结论与贡献定位

This paper addresses the common issue of "generalization extrapolation" in discussions of learning methodologies. Methodologically, we distinguish between the acquisition phase (moving from inability to ability) and the internalization phase (moving from ability to stable invocation). This research focuses specifically on the critical task of the internalization phase: whether a learner can initiate and advance the next response path within an acceptable time window given specific situational cues (cue-response path mapping). On this basis, the paper proposes the Micro-Bottleneck Internalization Protocol (MBIP). The objective is to provide learners with an operational framework for self-scheduling that is repeatable, capable of recording process signals, and verifiable, rather than offering a set of empirical tips claiming universal cross-task applicability.

The core logic of MBIP is to converge practice from "full-problem/full-paragraph reproduction" toward the microscopic "stuck points" that truly consume time and determine performance, using a fixed procedure to repeatedly hit these bottlenecks. The protocol consists of three key components and sequential constraints: First, the i-Zone is used to locate recurring "error/slow/stagnation" bottleneck signals in a task, prioritizing the allocation of practice resources to the micro-links most in need of repair. Second, the Minimum Scenario (MS) encapsulates the bottleneck into a reproducible trigger prompt, following the principle of "just enough cues" so that the learner enters the specific local context and attempts to initiate the next step upon seeing the prompt. Third, at the point of retrieval failure, a Scenario-based Self-Explanation (SBE) is generated and solidified. The intended audience for this explanation is "one' s future self" ; thus, the explanation can be compressed into a minimum necessary scaffold, focusing on "how to initiate the next step here" or "what the key rationale is," and saved as reusable material via voice or short text. Procedurally, MBIP enforces a strict "retrieval first, correction second" sequence: the prompt only triggers retrieval, and explanatory materials are invoked only after the attempt, preventing the act of "reading the explanation first" from replacing active recall and masking failure signals.

To prevent long-term drift during fragmented training, MBIP further proposes a dual-track system consisting of a "Micro-Reinforcement Track" and a "Macro-Structural Track." The micro-track executes high-frequency retrieval and feedback loops around the i-Zone, while the macro-track maintains the retrieval of antecedent relationships and overall structure through outlines or concept maps, requiring each MS to have a clear index position within the macro-structure. When a "retrieval + invocation of explanation" fails to advance progress, a structural diversion is prioritized: micro-practice is paused to return to the macro-structure for link repair before returning to the MS to continue internalization. Consequently, MBIP integrates "bottleneck localization, triggered retrieval, scaffold generation/invocation, and failure backtracking" into a repeatable closed loop. It describes operational states using process signals closer to actual practice (e.g., whether initiation is smoother, whether "stuck" moments are reduced, whether dependence on explanations is decreasing, and whether structural positioning is more stable) to support subsequent research design and evidence alignment.

Furthermore, this paper notes that MBIP is theoretically independent of specific tools or media formats; its core lies in reproducible training units and procedural constraints. Modern digital carriers (especially audio recording and playback) serve merely as implementation enhancements to reduce the friction of storage, invocation, and mobile reproduction. The mathematical case study in the appendix is intended to demonstrate one specific implementation path of the protocol, from bottleneck localization to transfer verification.

Agarwal P K, Nunes L D, Blunt J R. Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms[J]. Educational Psychology Review, 2021, 33(4): 1409-1453. DOI:10.1007/s10648-021-09595-9.

Anderson J R, Corbett A T, Koedinger K R, et al. Cognitive tutors: Lessons learned[J]. Journal of the Learning Sciences, 1995, 4(2): 167-207. DOI:10.1207/s15327809jls0402_2.

Bisra K, Liu Q, Nesbit J C, et al. Inducing self-explanation: A meta-analysis[J]. Educational Psychology Review, 2018, 30(3): 703-725. DOI:10.1007/s10648-018-9434-x.

Boud D, Lawson R, Thompson D G. The calibration of student judgement through self-assessment:

Disruptive effects of assessment patterns[J]. Higher Education Research & Development, 2015, 34(1): 45-59. DOI:10.1080/07294360.2014.934328.

Cepeda N J, Pashler H, Vul E, et al. Distributed practice in verbal recall tasks: A review and quantitative synthesis[J]. Psychological Bulletin, 2006, 132(3): 354-380.

DOI:10.1037/0033-2909.132.3.354. Cepeda N J, Vul E, Rohrer D, et al. Spacing effects in learning: A temporal ridgeline of optimal retention[J]. Psychological Science, 2008, 19(11): 1095-1102.

DOI:10.1111/j.1467-9280.2008.02209.x. Chang S, Liu D, Tian R, et al. Automated procedural analysis via video-language models for AI-assisted nursing skills assessment[EB/OL]. arXiv:2509.16810, 2025. DOI:10.48550/arxiv.2509.16810.

Chi M T H, de Leeuw N, Chiu M H, et al. Eliciting self-explanations improves understanding[J].

Cognitive Science, 1994, 18(3): 439-477. DOI:10.1207/s15516709cog1803_.

Dunlosky J, Rawson K A, Marsh E J, et al. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology[J]. Psychological Science in the Public Interest, 2013, 14(1): 4-58. DOI:10.1177/1529100612453266. [10] Hattie J, Donoghue G. Learning strategies: A synthesis and conceptual model[J]. NPJ Science of Learning, 2016, 1. DOI:10.1038/npjscilearn.2016.13. [11] Karpicke J D, Blunt J R. Retrieval practice produces more learning than elaborative studying with concept mapping[J]. Science, 2011, 331(6018): 772-775. DOI:10.1126/science.1199327. [12] Karpicke J D, Blunt J R, Smith M A. Retrieval-based learning: Positive effects of retrieval practice in elementary school children[J]. Frontiers in Psychology, 2016, 7: 350.

DOI:10.3389/fpsyg.2016.00350. [13] Kobayashi K. Learning by preparing-to-teach and teaching: A meta-analysis[J]. Japanese Psychological Research, 2019, 61(3): 192-203. DOI:10.1111/jpr.12221. [14] Osterhage J L, Usher E, Douin T A, et al. Opportunities for self-evaluation increase student calibration in an introductory biology course[J]. CBE—Life Sciences Education, 2019, 18(2): ar22.

DOI:10.1187/cbe.18-08-0160. [15] Pan S C, Carpenter S K. Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice[J]. Educational Psychology Review, 2023, 35(4): 97. DOI:10.1007/s10648-023-09814-5. [16] Panadero E. A review of self-regulated learning: Six models and four directions for research[J].

Frontiers in Psychology, 2017, 8. DOI:10.3389/fpsyg.2017.00422. [17] Roediger H L , Karpicke J D. Test-enhanced learning: Taking memory tests improves long-term retention[J]. Psychological Science, 2006, 17(3): 249-255. DOI:10.1111/j.1467-9280.2006.01693.x. [18] Roediger H L, Butler A C. The critical role of retrieval practice in long-term retention[J]. Trends in Cognitive Sciences, 2011, 15(1): 20-27. [19] Seufert T. The interplay between self-regulation in learning and cognitive load[J]. Educational Research Review, 2018. DOI:10.1016/j.edurev.2018.03.004. [20] Shoesmith A, Hall A, Wolfenden

L, et al. Barriers and facilitators influencing the sustainment of health behaviour interventions in schools and childcare services: A systematic review[J].

Implementation Science, 2021, 16. DOI:10.1186/s13012-021-01134-y. [21] Shuell T. Phases of meaningful learning[J]. Review of Educational Research, 1990, 60: 531-547.

DOI:10.3102/00346543060004531. [22] Sweller J. Cognitive load theory[M]. In: Psychology of Learning and Motivation. Vol.55. Elsevier, 2011: 37-76. [23] Tenison C, Anderson J. Modeling the distinct phases of skill acquisition[J]. Journal of Experimental Psychology: Learning, Memory, and Cognition, 2016, 42(5): 749-767. DOI:10.1037/xlm0000204. [24] Tenison C, Fincham J, Anderson J. Phases of learning: How skill acquisition impacts cognitive processing[J]. Cognitive Psychology, 2016, 87: 1-28. DOI:10.1016/j.cogpsych.2016.03.001. [25] Vargas E A, Vargas J S. Programmed instruction: What it is and how to do it[J]. Journal of Behavioral Education, 1991, 1(2): 235-251. [26] Vygotsky L S. Mind in society: The development of higher psychological processes[M]. Cole M, John-Steiner V, Scribner S, et al, eds. Harvard University Press, 1978.

This appendix uses a single problem-solving and review session from Learner A' s execution example to illustrate how the Micro-Bottleneck Internalization Protocol (MBIP) handles encapsulation, field (micro-bottleneck) localization, and the execution of the Minimum Scenario (MS) within specific tasks. The example covers:

The operational steps of Minimum Scenario Self-Explanation (MS-SBE), enabling readers to replicate these steps in their own learning. It must be emphasized that MBIP describes a set of processes and constraints rather than a specific software tool; the same workflow can be implemented across various platforms, such as flashcard software or note-taking systems. This includes the pre-setting of MS during the card-creation phase, the anchoring of macrostructures after creation, the MS-SBE process during spaced repetition, and the protocols for failure branching and structural backtracking. The objective is to allow the reader to paraphrase the MBIP logic and reuse it based solely on this description.

Task Context and Localization: Given vectors $\mathbf{a} = (2, -1)$, $\mathbf{b} = (3, \lambda)$, and $\mathbf{c} = (2, 4)$, if $2\mathbf{a} - \mathbf{b}$ is collinear with $\mathbf{b} + 3\mathbf{c}$, find λ. Learner A is capable of performing basic vector algebraic operations. For instance, A can calculate: $2\mathbf{a} - \mathbf{b} = (1, -2 - \lambda)$ and $\mathbf{b} + 3\mathbf{c} = (9, \lambda + 12)$. However, a stagnation occurs at the next step: when the problem provides the "collinear/parallel" condition, A cannot stably translate this natural language cue into a computable mathematical criterion (hesitating between proportional relationships and determinant-based criteria, or repeatedly rewriting incorrect equations). According to the MBIP performance signal standards (Error/Slow/Stagnation), the i-zone in this example is localized as: The rapid translation of the cue "two vectors are collinear/parallel" into a mathematical equation to initiate the solution (Cue-Response Path Mapping).

The card-creation principle of MBIP dictates that the intervention unit is not the playback of the entire problem, but rather the "translation and initiation action" localized within the i-zone, encapsulated into a reproducible card. The card structure consists of "Question Side (MS) + Answer Side (SBE)." In this example, Learner A employs "dual audio" to reduce interaction costs during mobile review: the front-side audio reads the trigger cue, while the back-side audio provides a minimalist, verbalized correction. The Question Side (MS) contains the minimum trigger prompt and necessary simplifications: the MS retains only the minimum set of cues required to trigger the i-zone, thereby avoiding the dilution of training intensity caused by redundant background information.

1.2.1 MS

In this example, the bottleneck does not lie in the vector operations themselves, but rather in the "mathematical expression of the collinearity condition." Therefore, the Memory Stimulus (MS) abstracts the vectors from the specific problem into a general form, focusing attention on the translation action of "collinearity → equality." Learner A adopts the following MS: $\vec{u} = (x_1, y_1), \vec{v} = (x_2, y_2)$.

MS (Front Text): Given vectors $\vec{u} = (x_1, y_1)$ and $\vec{v} = (x_2, y_2)$. It is important to clarify that an MS is not merely a "shortened version of the original problem," but rather a minimal trigger prompt directed at the Zone of Proximal Interference (i-zone). Its goal is to use the least amount of information possible to consistently reproduce the same bottleneck reaction (the same "clue-response path mapping"). Whether such a simplification is successful can be judged by an operational criterion: when the learner sees the MS, do they still encounter the same "translation threshold" as in the original problem and trigger the same type of hesitation, error, or activation path?

If the simplification fails to trigger the bottleneck, it indicates that there are too few clues and information must be backfilled; conversely, if the MS still contains a large amount of irrelevant information, it indicates that excessive clues will dilute the training effect. To enhance learning outcomes, the MS can be configured with a "Front Trigger Audio" recorded by the learner themselves. During the card creation stage, the learner reads the trigger clue as a short 3–6 second audio clip. This allows them to enter the "pre-retrieval" step using only auditory cues during review (i.e., attempting to generate the next response before being exposed to the materials on the answer side). It must be emphasized that MBIP does not assume all MS are suitable for eyes-closed or purely auditory review. When an MS relies on critical visual information such as graphs, geometric configurations, or charts, the learner must still view the image to complete the retrieval; in such cases, the audio serves only as a supplementary prompt rather than the primary clue. In this specific example (which does not rely on images), Learner A recorded the following trigger audio during card creation:

Front Audio (Trigger Audio, recommended 3–6 seconds): "Given that vectors

$\vec{u}$ and $\vec{v}$ are collinear/parallel, what equation do they satisfy?”

(Answer Side): Minimalist formulaic explanation.

1.2.2 SBE

SBE (Self-Generated Brief Explanation) is generated and recorded by the learner during the card creation phase, serving as a fixed explanatory scaffold on the answer side. To reduce the cognitive burden of review and increase initiation speed, Learner A provides the shortest possible verbal expression of the directly applicable equation. This allows for the rapid recovery of the subsequent response path under similar Micro-Situations (MS).

SBE (Back-side audio, recommended duration 3–8 seconds): “If u and v are collinear, then⋯” The first step after card creation is “Macro-Anchoring” : linking the card’ s key points (text and images) to a mind map or review outline (the macro-structure track). Once the card is completed, Learner A first executes this anchoring by copying the core textual content of the card—specifically the front-side MS text and the back-side SBE text/command—along with any images directly related to retrieval (essential if the card relies on visual information) to the corresponding node in the mind map or review outline (e.g., the “Vector Collinearity/Parallelism Criteria” node). This establishes the card as a reproducible training unit under that specific node.

Audio files (front-side trigger audio and back-side command audio) are typically stored within the flashcard system or as local files. They are linked to the node via filenames, hyperlinks, or tags, rather than requiring a full transcription of the script within the outline itself.

This step ensures a consistent mapping between micro-level training and macro-level structure. It allows the learner to perform high-frequency retrieval at the card level while simultaneously locating the knowledge and its prerequisites within the broader structure. In the event of a “structural link failure,” the learner can quickly trace back to upstream nodes for remediation.

Review: Spaced Repetition (executable via flashcard software or other tools). After completing card creation and macro-anchoring, Learner A reviews the card according to a spaced repetition schedule. Each review session follows a standardized flow: (1) Present the MS (front-side text and images; if audio is recorded, it plays automatically). In this example, the audio “Given that vectors u and v are collinear/parallel, what equation do they satisfy?” plays automatically. (2) Retrieval: Independently generate the mathematical conditions for “collinearity/parallelism” within a limited timeframe (e.g., 5–10 seconds).

In this case, successfully recalling the target equation clearly in one’ s mind is counted as a success. (3) Correction: After reflecting, the learner clicks to reveal the text or replay the answer-side SBE (back-side command audio or text) to verify and correct their response.

Note: During the review phase, the recording plays automatically when the back of the card is displayed. The learner can then select the subsequent review interval based on their level of mastery.

It should be added that for image-dependent cards (e.g., cards requiring the observation of geometric figures, function graphs, or apparatus structures to make a judgment), the retrieval phase generally cannot be completed through auditory input alone. The learner must view the image before attempting retrieval; in these instances, the audio serves primarily as a prompt and an auxiliary tool for correction rather than a substitute for visual input.

Furthermore, MBIP is not dependent on specific software. During the review phase described in Section 1.4, PowerPoint (PPT) can serve as a medium to implement the same process as flashcard software: (a) The "Front (MS)" and "Back (SBE)" can be placed on two adjacent slides, or a single slide can use step-by-step animations to ensure the front is seen before the back appears. (b) Audio files (front-side trigger and back-side command) can be embedded directly into the slides. (c) The review sequence can be customized by reordering the PPT slides (e.g., by chapter, prioritizing weak points, or following a daily plan) to achieve personalized scheduling. Additionally, using the "Auto-play" function in PPT allows for the hands-free review of multiple cards in succession.

Therefore, at the tool level, learners may choose an implementation method based on their devices and habits. At the process level, steps 1.1–

1.3 的定位、制卡与挂靠步骤保持不变。

Failure Shunting: Card Fine-tuning. When the same Memory Slot (MS) fails frequently across multiple retrieval sessions, the MBIP (Memory-Based Iterative Process) implements failure shunting through two distinct pathways. Path 1 (Local Fine-tuning) is applied if the learner knows "which equation should be written"but remains prone to confusion, or if they rely heavily on audio/text cues from the answer side to generate the response. In such cases, the process returns to the card-creation stage for refinement. This involves further compressing the "passphrase" on the back of the card or adding minimal prompts to the front (e.g., "write as a proportion") to reduce irrelevant hesitation and initiation friction. Alternatively, more comprehensive self-explanation may be added to the back of the card.

Path 2 (Structural Disconnection) is triggered if the learner cannot retrieve the prerequisite concepts for "proportionality," rendering them unable to understand or replicate the content even after reviewing the passphrase. In this scenario, the learner is redirected to the macro-structural track to supplement missing prerequisite nodes (definitions, equivalent conditions, and minimal illustrative examples). Once these structural gaps are repaired, the learner returns to the original card to continue the retrieval process.

Figure 1

Figure 1: Figure 1

Figure 2

Figure 2: Figure 2

1.6 回到原题：以迁移验证 i 区是否被内化

Once the agent can stably and rapidly generate collinearity criteria within the Mental Space (MS), it can return to the original problem to complete the solution smoothly. Specifically, after calculating the coordinates of the two vectors, the agent immediately establishes the collinearity condition and solves for the unknown. From the perspective of the Model-Based Internalization Process (MBIP), the key to internalization is not merely "being able to solve this specific problem," but rather ensuring that when "collinearity/parallelism" cues reappear in future tasks, the learner can more quickly and stably activate the same translation and solution pathway.

Figures

Source: ChinaXiv —Machine translation. Verify with original.

Figure 3

Figure 3: Figure 3

Figure 4

Figure 4: Figure 4

Figure 5

Figure 5: Figure 5