

A Systematic Review of the Comprehension and Retention Level of STEM Subjects among Secondary School Students

Maria Christoforaki, Athina Karatza, Myrto Koutra-Illiopoulou*, Anastasia Georgiou, Nelly Marosi, Eirini Chatzara, Evangelia Mavrikaki, Apostolia Galani

Department of Pedagogy and Primary Education, National and Kapodistrian University of Athens, Athens, Greece

*Corresponding author: lialani@primedu.uoa.gr

Received April 23, 2026; Revised May 25, 2026; Accepted June 01, 2026

Abstract This systematic review comprises a synthesis of 204 empirical studies published in 2025, examining how secondary school students comprehend and retain knowledge across STEM disciplines. The analysis of the selected articles concerning comprehension was based on an interpretative framework, distinguishing between knowing why (conceptual) and knowing how (procedural) understanding. The review identified asymmetries among students' comprehension. While students exhibit strong procedural fluency, their conceptual reasoning appears to be fragmented. Regarding retention, only four studies examined it, indicating that procedural knowledge persisted while conceptual understanding decayed in the absence of iterative reflection. Most of the identified difficulties were clustered around STEM topics such as energy transformations, chemical bonding, cellular respiration, and mathematical abstraction, underscoring conceptual fragility in representationally complex domains. The synthesis calls for longitudinal research to trace meaningful STEM learning over time to identify the ideal practices that promote iterative and reflective reconstruction of comprehension and retention.

Keywords: *Comprehension, Retention, STEM subjects, Secondary education*

Cite This Article: Maria Christoforaki, Athina Karatza, Myrto Koutra-Illiopoulou, Anastasia Georgiou, Nelly Marosi, Eirini Chatzara, Evangelia Mavrikaki, and Apostolia Galani, "A Systematic Review of the Comprehension and Retention level of STEM Subjects among Secondary School Students." *American Journal of Educational Research*, vol. 14, no. 5 (2026): 149-158. doi: 10.12691/education-14-5-4.

1. Introduction

In recent years Science, Technology, Engineering and Mathematics (STEM) education has been spotted at the forefront of the curricula reform initiatives, as it serves a wide spectrum of objectives including the enrichment of students' knowledge and skills, while addressing the enactment of goals related to equity and inclusion [1]. Particularly, the development of secondary-school students' comprehension and retention regarding STEM have become a sustained concern in educational research, arising from the limitation of STEM learning to the acquisition of facts and the accurate execution of procedures.

The level of cognitive development of secondary school students represents a pivotal stage in cognitive and identity formation, when learners transition from concrete to formal operational thinking [2]. Hence, they are expected to interpret abstract concepts, use symbolic and visual representations, connect ideas across domains, and apply their knowledge to unfamiliar problems, shaping their future engagement with STEM fields [3].

Moreover, international assessments such as the

Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) consistently highlight this period as critical for developing foundational STEM literacy and for mitigating widening gaps in conceptual understanding [4,5]. Therefore, a central challenge is the explanation of secondary school students' progress from procedural forms of knowledge, towards a vertically coherent comprehension and retention of the obtained knowledge and skills in STEM disciplines.

This challenge has been addressed from a wide range of theoretical frameworks, including conceptual change theory [6,7], examining the cognitive dissonance of the prior conceptions of students to the scientifically sound explanations, and inquiry-based learning [8,34] highlighting the student-centered investigation in helping learners construct meaning through engagement with the scientific phenomena. Supplementary to the abovementioned is the encouragement of students' reasoning across disciplinary boundaries, overcoming the distinct separation of school subjects [9,10].

Each of these frameworks foregrounds a different dimension of STEM learning, expanding the conceptualization of STEM learning beyond factual recall and procedural performance. Therefore, the unifying goal

is to understand both the comprehension mechanisms of learners regarding core STEM ideas and the practices that support the retention of knowledge understanding. However, empirical evidence reveals variations in students' grasp of fundamental STEM concepts, as well as persistent difficulties in retaining and applying this knowledge over time [5,11].

Within this body of work, while the construct of comprehension is commonly used, it is not approached with systemic precision. During our study, we have approached the concept of students' comprehension, as an integrated understanding of STEM content, embracing two aspects: (a) conceptual knowledge, regarding the reasoning of occurrence of the scientific phenomena, (b) procedural knowledge, concerning the performance of the procedures, along with the extent to which knowledge is transferred, maintained and reapplied across both contexts and temporal intervals [12].

We acknowledge that contemporary theoretical accounts of STEM learning foreground discrete content constellation of a broader repertoire of epistemic and cognitive practices that includes dimensions such as analytical reasoning, modelling practices, metacognitive self-regulation, interdisciplinary synthesis, and engagement with socio-scientific issues [12,13,14]. However, the present study circumscribes its attention to the conceptual and procedural dimensions of comprehension. This delimitation is justified by the fact that these constructs are among the most consistently operationalized in the empirical literature, thereby allowing for a more coherent synthesis across heterogeneous studies.

Moreover, a broader socio-cultural turn in STEM learning indicates that is not only an individual cognitive achievement, but also a situated process, mediated by classroom practices, social interactions, prior experiences, and students' evolving identities as STEM learners [15,16,37]. From this vantage point, the construct of conceptual retention becomes central to understanding both the durability and coherence of students' STEM learning, making it another core focus of the present study, as this perspective demands to map (a) What remains?, (b) Why do some STEM concepts resist forgetting while others fade or become fragmented?

In parallel, a significant body of research has identified several STEM-topic difficulties that impede either the comprehension or the retention across the STEM disciplines. For example, [17] emphasize that the effective problem-solving in math discipline is based on students' ability to both construct and refine their conceptual models, while developing skills to construct mathematical proofs bears notable barriers [18]. On the contrary, students face persistent misconceptions regarding fundamental concepts of the physics and chemistry disciplines such as forces, energy transformations, and particle interactions, which considerably affect student learning [19,35]. These identifications demand the creation of clusters of the specific STEM-topics that students find most to capture the when conceptual fragility is most likely to emerge and persist retention of the STEM concepts and skills [20,21,36].

However, despite these valuable contributions, the field lacks a systematic synthesis that connects students' levels

of comprehension, their retention of STEM concepts over a designated period, and the specific topics that students find most difficult to grasp and retain, and that does so with secondary school students in focus. A systematic review of existing empirical research can reveal how students currently comprehend and retain STEM concepts, identify persistent conceptual barriers, and illuminate implications for curriculum design, instructional support, and policy. Accordingly, this study provides a comprehensive synthesis of the empirical literature addressing the following questions:

(R.Q.1) What is the current comprehension level in STEM subjects among secondary school students?

(R.Q.2) To what extent do secondary school students retain concepts of STEM subjects over a predetermined period?

(R.Q.3) Which specific topics of STEM subjects do students find most difficult to understand and retain?

2. Materials and Methods

2.1. Review Design

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 guidelines [22]. This strategy was selected to ensure transparency, while developing an in-depth understanding of the current comprehension and retention level of STEM subjects among secondary school students.

2.2. Database search and Search Terms

The databases used to retrieve articles, were based on the purpose of our review and its further determination, according to the research questions mentioned above, namely:(a) Scopus, (b) ERIC, (c) Web of Science, (d) Google Scholar and (e) PsycINFO.

The research covered publications from January 2025 to October 2025, reflecting the most recent research developments. As a first step to our searching strategy, we created keyword-strings relevant to our constructions of interest.

Table 1. Search strings used to screen the selected databases

Dimension	Search string
Comprehension level	"STEM" AND "education" OR (sciences OR technology OR engineering OR mathematics OR math) AND ("comprehension" OR "understanding" OR "concepts" OR misconceptions OR ideas) AND ("secondary students OR "secondary education" OR "secondary school students" OR "high school students" OR secondary)
Retention level	"STEM" AND ("STEM concept retention" OR "concept retention") AND "education" AND concepts AND ("secondary students" OR "secondary education" OR "secondary school students" OR "high school students" OR secondary)

The keyword-strings that we selected to use (Table 1) are based on the scope and focus of the selected database, considering them as an optimal strategy to achieve adequate coverage of our researching field. Nonetheless, it is important to clarify that no additional search was

required for R.Q.3, as it could be addressed based on the findings from the previous research questions (R.Q.1 and R.Q.2).

2.3. Inclusion and Exclusion Criteria

To be included in this review, a study should endorse the following criteria:

- The study must focus on the understanding related to both the conceptual (know why) and procedural comprehension dimension (know how) of the STEM subjects.
- The study must perform empirical research reporting quantitative, qualitative, or mixed-methods data.
- The context of the study must be situated within secondary education.
- The study should be published in a highly ranked peer-reviewed journal in 2025, according to citation scores and impact factors of 2025.
- It should be available in English.

Whereas, a study was excluded from the database, when:

- Addressing STEM attitudes, motivation, or interest without measuring comprehension or retention.
- Focusing on gender or equity issues under STEM education without measuring comprehension or retention.
- Regarding conceptual and theoretical papers without empirical data.
- Containing primary research results or being at an early stage of investigation.
- Including conference abstracts, editorials, or non-peer-reviewed sources.
- Being only partially available or inaccessible to us.
- Failing to meet the quality research standards (short papers, e.g. short conference papers with limited methodological detail) according to the CASP checklist.

2.4. Selection Process

We conducted the selection process of the yielded articles, according to the three distinct stages implied by the PRISMA guidelines (Figure 1). Firstly, during the title and abstract screening, all retrieved records were independently screened by two reviewers to remove duplicates and obviously irrelevant studies. Secondly, we reviewed the full text articles, according to the inclusion/exclusion criteria. In every case the two reviewers identified discrepancies, they resolved them discussion and until they conclude to a consensus. Thirdly, the final inclusion included all the studies meeting all the inclusion criteria determined for this review. We should clarify that during the full-text review, the Critical Appraisal Skills Programme (CASP) checklist 1 was applied to evaluate both the methodological quality and validity of each study, ensuring that only studies with sufficient rigor were considered for inclusion.

As illustrated in Figure 1, during the initial search we identified 1,175 records across the selected databases. Of these, 500 duplicate records were removed, and an additional 100 records were marked as ineligible by automation tools, resulting in 575 records to be screened by title and abstract. The screening stage led to the exclusion of 250 records that did not meet the inclusion criteria of the study. The remaining 325 articles were retrieved for full-text review and assessed for eligibility according both to the inclusion/exclusion criteria and the CASP checklist. During this stage, 121 reports were excluded with explicit reasons. In particular:

- 40 articles concerning non-empirical research.
- 31 articles addressing STEM attitudes, motivation, or interest without measuring comprehension or retention.
- 50 articles focusing on gender or equity issues under STEM education without measuring comprehension or retention.

In total, 204 studies were chosen to be analyzed for the final synthesis.

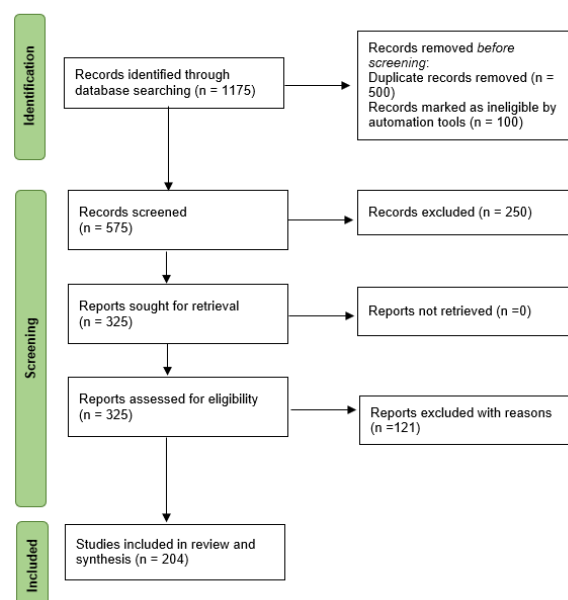


Figure 1. Flow chart showing the stages of the systematic review

2.5. Conducting the Review

2.5.1. Analysis Approach

Each of the 204 articles was analyzed following a two-tier analytical procedure, in which the first level functioned as a foundational and supplementary stage for the subsequent level of analysis. Specifically, the first tier entailed the initial phase of a detailed data collection, while the second tier followed thematic analysis to identify patterns and constructing interpretive themes, regarding the comprehension and retention level, along with the identified STEM topics that students face most of the cognitive obstacles. This strategy led to a nuanced understanding of how evaluation practices were conceptualized and implemented across the reviewed studies. Regarding the initial phase of the data collection, we extracted and organized the data using an Excel

1 The CASP checklists are available from the Critical Appraisal Skills Programme official website: <https://casp-uk.net/casp-tools-checklists/>

worksheet, which functioned as a structured tool for managing the information. The process of extraction was unfolded in three distinct stages.

Stage 1: We created five main sections of the worksheet, each including columns for relevant sub-categories. The first section captured the basic bibliographic information of the articles, including: (a) country, (b) title, (c) authors' details, and (d) the database that was retrieved. The second section entailed 4 columns of the STEM subjects (science, mathematics, technology, engineering), along with two columns recording the methodology design (sub-categorized into experiments, quasi-experiments, and non-experiments), the types of data (sub-categorized into quantitative, qualitative, and mixed methods) and the instruments (questionnaires, interviews, worksheets, observation rubrics). The remaining three sections were designed to record information corresponding to the three research questions of the present review (Table 2). Particularly:

(R.Q.1): We created three columns to record the three comprehension tiers (basic, developing, integrated). The categorization of the results of each paper in relation to the three comprehension tiers is detailed in Stage 2.

(R.Q.2): We devoted three columns to capture the construct variables being evaluated and the determined retention focus examined for each article across the STEM subjects.

(R.Q.3): We concluded the worksheet with four columns recording the specific STEM topics that students find most difficult to retain and understand for each STEM discipline separately.

Table 2. Details of the analysis worksheet

Section	Columns / Sub-Categories	Purpose
Section 1	Country, Title, Authors' Details, Source	Capture basic bibliographic information of the articles
Section 2	STEM disciplines/ methodology design, types of data, instruments	Record STEM disciplines addressed in each study/ Mapping the research methodology
Section 3 (R.Q.1)	The three comprehension tiers (basic, developing, integrated)	Document the comprehension tier that its papers results were categorized.
Section 4 (R.Q.2)	Construct Variable(s), Retention Period	Capture the variables evaluated in each study and the retention focus examined across STEM disciplines
Section 5 (R.Q.3)	STEM Topics	Identify the specific STEM topics that students struggle with in terms of retention and understanding

Stage 2: As already mentioned, in this review, we approach the constructs of comprehension and retention regarding both the conceptual (why) and the procedural (how) aspect of them. Given this direction, we created two variables to provide a coherent and theoretically informed categorization of the findings: (a) Current Comprehension Level (C_C_L) and (b) Retention Level (R_L), examining each variable across two analytical dimensions knowing/retaining why and knowing/retaining how. Through this dual framework, we captured the multifaced nature of students' understanding within STEM disciplines, enabling a systematic investigation of the

ways the concepts of comprehension and retention are mapped in the literature.

We should clarify that through the presenting study the notion of level does not denote a standardized or numerical metric, but a qualitative indicator of the depth, coherence, and character of students' understanding as articulated within the distinct studies. As the evidence base proved to be rather heterogenous, direct comparison of aggregation of data was neither possible nor desirable. Instead, we preferred a narrative synthesis [23] of the results after identifying the orientation of each study between comprehension and retention, in relation to the conceptual or the procedural aspect.

This interpretive strategy enabled us to foreground patterns of convergence and divergence across the studies, while emphasizing how different methodological traditions have conceptualized and operationalized students' comprehension and its retention over the time.

To further enhance interpretive coherence and analytic transparency, we employed an inductive three-tier framework to categorize students' comprehension across the reviewed studies. This framework does not imply a standardized hierarchy or quantifiable scale, but rather a qualitative continuum of meaning representing how understanding was described, evidenced, and interpreted within each article. The developed framework enabled us to identify patterns of conceptual and procedural comprehension consistently across the heterogeneous datasets, inductively through iterative coding of the reported outcomes and theoretical framings.

The framework was layered into three tiers: (a) basic, (b) developing, and (c) integrated comprehension to capture the progressive degrees of the conceptual depth, coherence, and transferability (Table 3). Each category reflects distinct epistemic characteristics and indicators of reasoning, derived from how comprehension was operationalized in the primary literature.

Table 3. The three-tier interpretive framework for categorizing students' current comprehension in STEM

Comprehension tier	Indicators
Basic	Responses rely on memorization or direct substitution of formulas. Correct procedural outcomes with persistent misconceptions. Absence of causal or relational explanations. Low evidence of transfer or metacognition.
Developing	Explanations combine correct and naïve reasoning. Conceptual success in familiar tasks but failure in novel ones. Early evidence of reflection or awareness of reasoning gaps. Procedural fluency begins to support conceptual thinking.
Integrated	Construction and critique of models, arguments, or designs. Demonstrated transfer to new or complex problems. Explicit causal and relational explanations. Evidence of metacognition and abstraction.

Basic comprehension

The first tier of basic comprehension refers to the surface-level understanding characterized by the recall or

the reproduction of isolated facts, definitions, or algorithms. Students categorized at this level can state or execute correct responses but show limited evidence of reasoning beyond immediate recall. During this tier, conceptual connections remain fragmented or absent, while procedural competence often manifests as rote algorithmic execution rather than understanding of underlying principles.

Developing comprehension

The second tier of developing comprehension denotes partial conceptual integration, where learners begin to connect ideas or methods but their reasoning remains emergent and unstable. Central observation of this tier is that understanding appears inconsistency across contexts, as students may articulate partial causal explanations or apply correct procedures with incomplete conceptual grounding. This category reflects a transitional phase between reproduction and integration. Knowledge structures are undergoing reconstruction, while procedural fluency begins to support conceptual understanding. However, coherence has not yet achieved.

Integrated comprehension

The tier of integrated comprehension represents a coherent and transferable understanding in which both the conceptual and procedural knowledge are mutually reinforcing. Learners at this level can explain, justify, and apply their understanding across novel or complex contexts. Comprehension is no longer confined to recall or routine application but involves flexible reasoning, abstraction, and metacognitive reflection. Particularly, this tier embodies the conceptual coherence and the procedural fluency as a single epistemic act, representing a disciplinary understanding that is both deeply grounded and generative, enabling learners to reconstruct knowledge in response to the new challenges.

This three-tier framework should not be interpreted as a rigid hierarchy, but as a dynamic continuum of cognitive and epistemic development, bridging cognitive and sociocultural perspectives (Bransford et al., 2000; Nasir, 2012).

Stage 3: To address the third research question, we performed a secondary-level analysis, integrating the evidence extracted from both RQ1 (current comprehension level) and RQ2 (retention level). Through this secondary analysis, we aimed to identify the specific STEM topics that consistently emerged as areas of persistent student difficulty, either in terms of conceptual understanding (knowing why), procedural fluency (knowing how), or the long-term retention of both.

To do so, we followed the principles of convergent narrative synthesis [23] where we re-examined the topic-level data across the reviewed studies. Each study's focal content area was coded according to: (a) STEM discipline and subfield (e.g., algebra, forces, chemical bonding), and (b) the dimension of difficulty explicitly reported or implied. These dimensions were then mapped to the interpretive framework established in this review (Table 3), conceptual (knowing why) and procedural (knowing how), to trace patterns of comprehension fragility and retention decay.

The integration of both data streams enabled us to construct a difficulty continuum and link conceptual fragmentation, procedural instability, and retention

vulnerability. It is important to highlight that this interpretive process did not involve new quantitative aggregation, but rather a systematic cross-coding and clustering of the findings, preserving both the methodological integrity and the contextual richness of each study.

3. Results

3.1. Overview of Included Studies

As shown in Table 4, the studies included in this review were drawn from a diverse range of high-impact journals in the field of science and STEM education. The International Journal of Science Education (n = 37) and the Journal of Research in Science Teaching (n = 31) accounted for the largest proportion of the publications, reflecting their longstanding contribution to empirical research on students' learning and understanding in STEM contexts. Other key outlets such as Science Education (n = 26) and Research in Science Education (n = 22) also featured prominently, underscoring the strong representation of science education scholarship within the corpus.

Notably, a growing number of studies were published in explicitly STEM-oriented journals such as the International Journal of STEM Education (n = 21) and the European Journal of STEM Education (n = 17), indicating the field's gradual movement toward integrative and interdisciplinary perspectives.

Table 4. Distribution of the studies by journal

Journal	Number of Studies (n)
International Journal of Science Education	37
Journal of Research in Science Teaching	31
Science Education	20
Research in Science Education	22
International Journal of STEM Education	21
European Journal of STEM Education	17
Journal of Science Education and Technology	14
Journal of Science Teacher Education	14
School Science and Mathematics	13
Journal of Research in STEM Education	11
STEM Education	1
Educational Studies	1
Interactive Learning Environments	1
PLOS ONE	1

Among the above studies, as illustrated in Table 5, most studies employed quantitative designs (n = 105), reflecting a strong emphasis on measuring comprehension and retention through tests, concept inventories, and large-scale assessments. This trend was particularly evident in science (n = 46) and mathematics (n = 35) education, where standardized instruments and statistical models have been widely used to examine conceptual understanding and long-term learning outcomes.

In contrast, qualitative studies (n = 61) were more prevalent in technology and integrated STEM contexts, often using classroom observations, interviews, or discourse analyses to capture students' reasoning processes and the situated nature of comprehension.

Mixed-methods research ($n = 46$) has gained increasing prominence, especially in integrated STEM and engineering education, as scholars have sought to combine cognitive measures with rich contextual insights into learning processes.

Table 5. Distribution of studies by STEM subject and research design

STEM subject	Total studies (n)	Quantitative (n)	Qualitative (n)	Mixed methods (n)
Science (Physics, Chemistry, Biology, Earth Science)	80	43	23	14
Mathematics	62	35	15	12
Technology	22	8	9	5
Engineering	15	6	4	5
Integrated STEM (cross-disciplinary)	25	10	8	7
Total	204	102	59	43

Beyond the methodological tendencies observed across disciplines, it was also essential to examine how the reviewed studies positioned their focus in relation to the constructs of comprehension and retention. While methodological diversity reflects the field's epistemological maturity, the thematic orientation of research provides deeper insight into which aspects of STEM learning have received sustained scholarly attention. Distinguishing between studies that explore students' comprehension level, their immediate understanding of concepts and procedures, and those that investigate retention level, the endurance and transfer of that understanding over time, offers a more nuanced perspective on how the literature conceptualizes learning continuity. Table 6 summarizes this distribution, highlighting a clear asymmetry between the extensive body of research on comprehension and the relatively limited attention devoted to knowledge retention within secondary STEM education.

Table 6. Distribution of studies by research focus: Comprehension vs. Retention

STEM subject	Current comprehension level (C_C_L)	Retention level (R_L)
Science (Physics, Chemistry, Biology, Earth Science)	78	2
Mathematics	61	1
Technology	22	—
Engineering	15	—
Integrated STEM	24	1
Total	200	4

3.2. Current Comprehension Level in STEM Subjects among Secondary Education Students

The synthesis of the 200 studies reviewed revealed a complex and uneven landscape of students' comprehension across the STEM disciplines. In accordance with the analytical framework guiding this review, comprehension was examined along two intersecting dimensions, knowing why (conceptual) and knowing how (procedural), and

categorized inductively into three qualitative tiers of understanding: basic, developing, and integrated comprehension. These tiers represent interpretive continua rather than standardized levels, allowing for nuanced differentiation in how studies articulated and evidenced students' understanding (Figure 2).

Across the full dataset, a clear pattern emerged: most studies clustered within the developing tier, followed by the basic tier, while instances of integrated comprehension were comparatively rare. This distribution was especially pronounced in the knowing how dimension, which dominated numerically across almost all domains. In contrast integrated comprehension appeared in far fewer studies, indicating that deep conceptual integration remains limited within current secondary STEM learning contexts.

3.2.1. Science

The science education corpus, which encompassed 78 studies, nearly half of the dataset, revealed a persistent imbalance between conceptual and procedural comprehension. Across subfields such as physics, chemistry, biology, and earth science, students frequently demonstrated procedural fluency in applying formulas, conducting experiments, and completing algorithmic or step-based tasks, yet their conceptual grasp of underlying mechanisms remained fragmented and unstable.

In physics, comprehension was concentrated at the developing tier (knowing why–developing, $n = 6$; knowing how–developing, $n = 8$), with smaller counts for integration (knowing why–integrated, $n = 2$; knowing how–integrated, $n = 4$). Chemistry followed a similar pattern (knowing why–developing, $n = 6$; knowing how–developing, $n = 6$; knowing why–integrated, $n = 1$; knowing how–integrated, $n = 2$). Biology reflected comparable trends (knowing why–developing, $n = 5$; knowing how–developing, $n = 5$; knowing why–integrated, $n = 1$; knowing how–integrated, $n = 3$). In earth science, developing comprehension remained predominant (knowing why–developing, $n = 3$; knowing how–developing, $n = 3$), while integrated levels were limited (knowing why–integrated, $n = 1$; knowing how–integrated, $n = 2$).

3.2.2. Mathematics

The mathematics domain, represented by 61 studies, similarly revealed an asymmetry between procedural mastery and conceptual depth. Across algebra, geometry, and calculus, students displayed strong fluency in executing algorithms and symbolic manipulations but weaker understanding of the abstract relationships underpinning these operations.

In algebra, knowing how–developing reached 6 studies (Figure 2), followed by knowing why–developing ($n = 6$). Geometry exhibited a similar pattern (knowing why–developing = 4; knowing how–developing = 6), while calculus reflected slightly more conceptual balance (knowing why–developing = 4; knowing how–developing = 7).

3.2.3. Technology and Engineering

Studies in technology ($n = 22$) and engineering ($n = 15$) presented a distinct, more practice-oriented profile of

comprehension. Here, understanding was often contextualized, emerging through engagement in design, coding, or modeling tasks rather than through explicit conceptual instruction.

In coding and programming, comprehension was broadly distributed (knowing why–developing = 2; knowing how–developing = 3). Design and fabrication showed limited basic-level understanding, with comparable representation across developing and integrated tiers (knowing why = 2 each; knowing how = 3 and 2, respectively). Mechanics and systems displayed a small but balanced pattern, with knowing why–integrated = 2 and knowing how–integrated = 2. Integrated STEM

The Integrated STEM domain (comprising interdisciplinary projects, modeling & simulation, and socio-scientific issues) included 24 coded instances and showed the highest proportion of integrated comprehension across all domains. Interdisciplinary projects reflected five coded instances, with three at the integrated tier (one conceptual and two procedural). Modeling & simulation contributed seven coded instances, four of which were integrated (two conceptual, two procedural). The strongest concentration occurred in socio-scientific issues (SSI), with nine of twelve instances at the integrated level (two conceptual, seven procedural). Across the Integrated STEM group, knowing how–integrated accounted for 11 studies, and knowing why–integrated for five, jointly representing the greatest depth of integration across both cognitive dimensions.

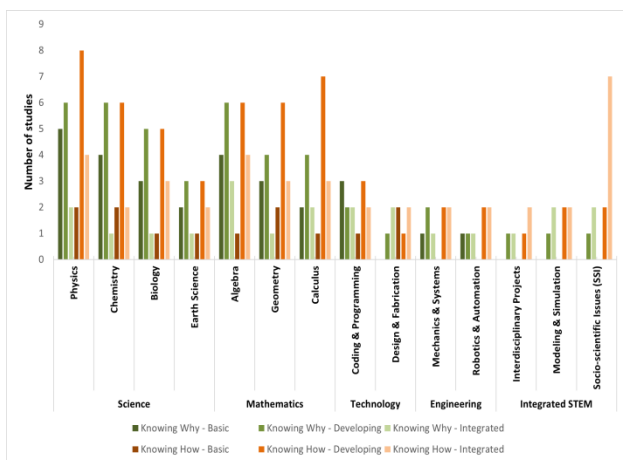


Figure 2. Qualitative synthesis matrix of comprehension levels across STEM domains

3.3. Current Retention Level in STEM Subjects among Secondary Education Students

Targeted searches were restricted to 2025 and only a small number of empirical studies met the inclusion criteria for the retention of STEM comprehension at the secondary level, specifically, those incorporating explicit delayed or post-delay assessments of conceptual and/or procedural understanding. This scarcity suggests that 2025 publications continue to privilege investigations of immediate post-instruction comprehension or adjacent constructs such as representational competence, modeling practices, and instructional design effects. In contrast,

longitudinal examinations of the durability of learning, including how conceptual and procedural knowledge are sustained or transformed over time, remain markedly underrepresented in the current literature.

Methodologically, this gap precludes the development of a same-year, within-journal synthesis of retention comparable in scope or depth to RQ1. Nevertheless, the few studies retrieved from adjacent journals point to critical directions for future inquiry: the need for systematic longitudinal research designs, cross-domain tracing of comprehension durability, and the integration of reflective and iterative assessment mechanisms capable of capturing learning as an ongoing, temporal process rather than a static outcome.

Although the number of 2025 studies explicitly addressing retention is limited, their collective findings offer a nuanced portrait of how students' conceptual and procedural knowledge evolve beyond immediate instruction. Each study employs delayed assessment to probe the temporal stability and transformation of understanding, illuminating which aspects of learning persist, which dissipate, and under which instructional conditions these trajectories occur.

3.3.1. Science

The study by [24] illustrates the temporal asymmetry between knowing how and knowing why. While students readily retained procedural fluency in calculating force and energy, their explanatory reasoning deteriorated in the absence of structured review. Spaced learning cycles, however, significantly mitigated this loss, underscoring that temporal distribution and re-engagement strengthen conceptual networks rather than merely rehearsing procedures. This aligns with the sociocultural framing of learning as reconstruction of meaning over time rather than the storage of information.

In chemistry education, [25] highlight the fragility of conceptual retention when learning remains decontextualized. Students maintained factual recall of hydrospheric processes yet lost explanatory reasoning about molecular interactions unless they participated in guided reflection. This echoes the long-standing critique that procedural competence, though durable, is insufficient for sustained understanding unless reconnected to conceptual frameworks.

3.3.2. Mathematics

[26] demonstrate how instructional design mediates retention. Their 5I-flipped model fostered durable comprehension through iteration, interaction, investigation, introspection, and integration, elements that required learners to continuously re-articulate mathematical relationships. The four-week delayed post-test revealed that students who engaged in reflective cycles retained both symbolic fluency and underlying conceptual coherence, whereas the control group exhibited sharp conceptual decay despite preserving algorithmic accuracy.

3.3.3. Integrated STEM

[27] extend this argument within an integrated, systems-thinking context. Their longitudinal design revealed that students who revisited carbon-cycle

interactions through modeling and discussion preserved integrated comprehension across cycles. Conversely, groups without reflective consolidation exhibited fragmented recall. The finding reinforces that retention thrives in iterative, meaning-making environments where students enact rather than merely recall knowledge.

Table 7. Empirical 2025 studies addressing retention of STEM comprehension

Domain	Subfield	References	Research approach	Determined period	Retention focus
Science	Physics	Zhou, Y., Hartley, R., Bernardelli, A., & Tolmie, A. (2025). The impact of spaced learning within physics lessons in secondary schools. <i>FLoS ONE</i> , 20(4), e021552. https://doi.org/10.1371/journal.pone.021552	Experimental (pre-post-delayed) N = 336 (UK)	7-day and ≈ 2-month delayed tests on energy and motion	Procedural problem-solving retained significantly better than conceptual explanations; spaced learning reduced decay of 'knowing why'.
Science	Chemistry	Co Oladejo, A. I., & Olatuju, T. T. (2025). Beyond the conventional flipped classroom: The 5I model in senior secondary school mathematics. <i>STEM Education</i> , 5(6), 974-999. https://doi.org/10.3934/steme.20250	Quasi-experimental (N ≈ 120, Nigeria)	4-week delayed post-test	Students in 5I-flipped model retained procedural skills and conceptual links longer than controls; reflection & iteration key factors.
Math	Algebra and Functions	Klemen, T., et al. (2025). Introduction of hydrosphere environmental problems in chemistry lessons through an online workshop. <i>Education Sciences</i> , 15(1), 57. https://doi.org/10.3390/educsci15010057	Pre-post-delayed (n = 110, lower secondary, Slovenia)	5-week delayed achievement test	Retention strongest for applied tasks; factual recall persisted; conceptual reasoning declined without follow-up reflection.
Integrated STEM	Earth science - Systems thinking	Mani, M., Palmberg, J., & Jernsen, E. (2025). An interactive learning environment for developing systems thinking on carbon cycling (Tracing Carbon). <i>Interactive Learning Environments</i> . https://doi.org/10.1080/10494830.2025.2494161	Design-based (longitudinal, Grades 7-9, Finland)	Multiple follow-ups across cycles	Sustained understanding of system interactions where reflection phases included; conceptual loss otherwise.

3.4. Topics of STEM Subjects Secondary Students Find Most Difficult to Understand and Retain

The synthesis of the 204 studies revealed recurring conceptual and procedural challenges clustered around specific disciplinary topics (Figure 3). To preserve analytical coherence, results were aggregated across comprehension dimensions, knowing why (conceptual) and knowing how (procedural), as well as the limited findings related to retention. This integrative approach emphasizes not isolated misunderstandings but enduring areas of conceptual fragility and procedural instability across STEM fields.

In physics, difficulties were most concentrated in electricity and circuits (n = 15), followed by force (n = 9), energy transformations (n = 9), and motion (n = 8). Across these topics, students often demonstrated procedural fluency in applying formulas or performing experiments yet exhibited shallow causal reasoning and incomplete explanatory models. Misconceptions about circuit behavior, force interactions, and energy transfer mechanisms were particularly persistent, suggesting limited integration between symbolic manipulation and underlying physical principles.

In chemistry, the most frequently reported area of difficulty was chemical bonding and molecular structure (n = 9), where symbolic fluency often masked conceptual misunderstanding. Additional challenges emerged in stoichiometry (n = 6) and conservation of matter (n = 3), where learners struggled to coordinate quantitative reasoning with particulate-level representations. Together, these findings highlight enduring gaps in students' ability to reconcile macroscopic observations with molecular-level explanations.

In biology, conceptual and procedural instability was observed primarily in topics of photosynthesis (n = 19) and cellular respiration (n = 6), with students recalling process sequences but failing to integrate them with underlying energetic and molecular dynamics. Persistent difficulties also appeared in genetics (n = 6) and inheritance patterns (n = 3), where understanding remained largely rote and fragmented rather than relational. These patterns indicate that learners often compartmentalize biological processes, limiting their capacity for systems-level reasoning.

In earth and environmental science, climate systems (n = 6), carbon cycle (n = 4), and plate tectonics (n = 3) were identified as recurring sources of conceptual confusion. Students encountered challenges in visualizing and reasoning about system dynamics over time, particularly when linking human activity to environmental feedback loops. Moreover, retention studies revealed that these understandings tended to decay without sustained engagement or cross-context reinforcement, as also reflected in RQ2.

Within mathematics, difficulties were most frequently reported in variables and functional relationships (n = 11) and spatial visualization (n = 11), followed by rate of change and accumulation (n = 9), proportional reasoning (n = 8), and proof and deductive reasoning (n = 8). Additional challenges were evident in limits and continuity (n = 5). Across these subfields, learners often demonstrated procedural accuracy but limited conceptual coherence, an asymmetry also reflected in the comprehension data for RQ1. The persistence of these patterns across decades of reform-oriented instruction underscores a structural tension in how mathematics mediates between symbolic proficiency and conceptual understanding.

In technology and engineering, conceptual difficulties mirrored those in science and mathematics. Topics such as logic and debugging (n = 6), optimization and system interdependence (n = 6), and abstraction and algorithm design (n = 3) revealed procedural competence without sufficient conceptual scaffolding. However, when representational transfer and systems integration (n = 8) and iterative design or reflection activities were incorporated, comprehension shifted toward integration, suggesting the importance of reflective cycles for stabilizing procedural understanding.

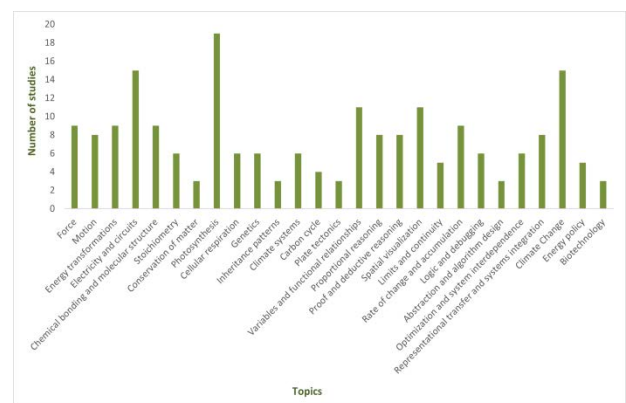


Figure 3. Specific STEM topics secondary students find most difficult to understand and retain

Finally, integrated STEM contexts, particularly climate change ($n = 15$), energy policy ($n = 5$), and biotechnology ($n = 3$), as well as interdisciplinary projects and systems modeling ($n = 6$) represented the most cognitively demanding yet potentially integrative learning environments. These topics required students to synthesize disciplinary knowledge with ethical reasoning and data interpretation. When adequately scaffolded, they fostered durable conceptual understanding and retention; when not, they exposed deep-seated fragility and cognitive overload.

4. Discussion

The synthesis of 204 studies reveals a complex and deeply instructive picture of how secondary school students comprehend and retain knowledge across STEM disciplines. Viewed through the dual analytical lens of knowing why and knowing how, understanding in STEM emerges not as a fixed endpoint but as a dynamic process of meaning-making that unfolds through interaction, representation, and time. Across domains, a consistent pattern appears: procedural fluency, the ability to execute tasks, apply formulas, or follow methodological steps, which tends to be stable and measurable, while conceptual coherence remains fragile, context-bound, and vulnerable to decay when reflection and integration are absent [28,29].

This asymmetry between procedural and conceptual understanding exposes a deeper epistemological tension in school STEM, where the reproduction of practice often overshadows the reconstruction of meaning. In science and mathematics, this imbalance is particularly evident. Learners demonstrate competence in problem-solving procedures yet struggle to articulate or justify the principles underlying those procedures [30]. In physics, students can compute forces, analyze motion, and quantify energy transformations, but they falter when connecting these operations to causal reasoning. In chemistry, they manipulate symbols and equations with proficiency yet remain uncertain about molecular interactions or conservation laws. In mathematics, learners execute algebraic manipulations or calculus operations accurately, but their grasp of structural relationships, limits, and rate of change is often superficial [31].

Such asymmetries, recurrent across decades and reaffirmed in the 2025 corpus, underscore a persistent disjunction between learning as procedural performance and learning as conceptual participation in disciplinary reasoning. The limited but illuminating body of retention-focused research further clarifies this relationship. Across studies in physics, chemistry, mathematics, and integrated STEM, comprehension was shown to endure not through repetition but through re-engagement and reinterpretation. Spaced learning cycles, reflective dialogue, and iterative modeling all supported the reconstruction of meaning over time [28,32].

Retention thus emerges as a fundamentally reconstructive process: understanding persists when learners revisit ideas in contexts that invite reinterpretation, rather than when knowledge is merely stored and rehearsed [31]. The three-tier interpretive framework applied in this review, basic, developing, and integrated

comprehension, captures this reconstructive trajectory. Students rarely progress linearly; instead, their understanding oscillates dynamically between partial coherence and integration as contexts shift. Basic comprehension reflects recall and task execution; developing comprehension involves emerging but unstable conceptual links; integrated comprehension entails the flexible coordination of conceptual and procedural knowledge across representations and contexts.

Progression between tiers depends on pedagogical conditions that support meta-conceptual reasoning, cross-representational translation, and opportunities for learners to articulate the why behind their procedural fluency [28]. Patterns of difficulty across disciplinary topics reinforce this interpretive depth. The areas most resistant to understanding, electricity and circuits, force and motion, energy transformations, chemical bonding and molecular structure, cellular respiration, photosynthesis, limits and continuity, and proof and deductive reasoning, share structural features that make them epistemic thresholds [18,25,33].

Each demands the integration of abstract representations with underlying mechanisms that are invisible, dynamic, or systemic. Students struggle not merely because these topics are complex but because traditional instruction isolates symbolic procedures from conceptual meaning. Even in integrated STEM contexts, including climate change, energy policy, and biotechnology, comprehension remains fragmented when reflection and modeling are absent. Conversely, when iterative design, systems thinking, and dialogic reasoning are embedded, learners achieve deeper, more durable understanding [29].

Viewed holistically, the corpus reveals a vertical pattern: students are most likely to retain what they have meaningfully integrated and most likely to forget what was learned procedurally without reflection. This insight aligns with constructivist and sociocultural perspectives that position learning as the reconstruction of meaning through participation in disciplinary practices. The scarcity of longitudinal and delayed-assessment studies in the field highlights an epistemological bias toward measuring immediate performance rather than sustained understanding [28].

Without deliberate attention to the temporal dimension of learning, research risks mistaking transient performance for durable comprehension. These findings collectively call for methodological, theoretical, and pedagogical recalibration. Methodologically, the field must expand beyond snapshot assessments to include longitudinal and mixed-methods designs capable of tracing how comprehension evolves, stabilizes, or erodes over time [30]. Theoretically, comprehension should be reconceived not as the sum of procedural and conceptual components but as their reciprocal interaction, each informing and transforming the other through cycles of practice and reflection.

Pedagogically, instructional design should emphasize recursive engagement through modeling, design-based inquiry, and spaced review, allowing learners to reinterpret procedures conceptually and test concepts procedurally. Reflection emerges as the mediating process that transforms experience into insight and stabilizes

understanding across time. Ultimately, comprehension in STEM should be understood not as a static achievement but as an ongoing reconstruction of meaning, a living process through which learners align experience, representation, and explanation. The balance between knowing how and knowing why cannot be restored by more content or practice alone; it requires learning environments that cultivate reflection, coherence, and transfer. Retention, in this sense, is not evidence of memory but of re-engagement: understanding endures because it continues to evolve. Designing for such enduring comprehension means embracing learning as a recursive act of making meaning again, where knowledge lives not in its repetition, but in its continual reconstruction.

ACKNOWLEDGEMENTS

The project and its research were undertaken within the scope of the Erasmus+ KA02 initiative under the acronym STEM-IT (Project Number: 21427).

References

- [1] European Commission. (2022). Relevant and high-quality higher education. Retrieved from <https://education.ec.europa.eu/kek/education-levels/higher-education/relevant-and-high-quality-higher-education>.
- [2] Piaget, J. (1972). *The psychology of the child*. Basic Books.
- [3] Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132.
- [4] Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 inter [26] results in mathematics and science. TIMSS & PIRLS Inter [26] Study Center, Boston College.
- [5] OECD. (2019). PISA 2018 results (Volume I): What students know and can do. OECD Publishing.
- [6] Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. *Inter [26] Handbook of Research on Conceptual Change*, 61–82.
- [7] Vosniadou, S. (2013). *Inter [26] handbook of research on conceptual change* (2nd ed.). Routledge.
- [8] Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107.
- [9] Beers, S. Z. (2011). *21st century skills: Preparing students for their future*. Pearson Education.
- [10] Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *Inter [26] Journal of STEM Education*, 3(1), 11.
- [11] National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. [26] Academies Press.
- [12] Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K–12 education: Status, prospects, and an agenda for research*. [26] Academies Press.
- [13] Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- [14] Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. NSTA Press.
- [15] Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- [16] Nasir, N. S. (2012). *Racialized identities: Race and achievement among African American youth*. Stanford University Press.
- [17] Lesh, R., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Lawrence Erlbaum Associates.
- [18] Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., ... & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, 23(7), 691–697.
- [19] Treagust, D. F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3(2), 297–328.
- [20] Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. [26] Academy Press.
- [21] Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the [26] Academy of Sciences*, 111(23), 8410–8415.
- [22] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097.
- [23] Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *ESRC Methods Programme*.
- [24] Zhou, Y., Hartley, R., Bernardelli, A., & Tolmie, A. (2025). The impact of spaced learning within physics lessons in secondary schools. *PLOS ONE*, 20(4), e0321552.
- [25] Klehm, W. D., Lichty, E., & Metellus, M. L. (2025). Improved student comprehension through student-designed tensile testing laboratory. *Proceedings of the 2025 ASEE Annual Conference*. <https://peer.asee.org/56756.pdf>.
- [26] Oladejo, A. I., & Olateju, T. T. (2025). Beyond the conventional flipped classroom: The 5I model in senior secondary school mathematics. *STEM Education*, 5(6), 974–999.
- [27] Mani, M., Palmberg, I., & Jeronen, E. (2025). An interactive learning environment for developing systems thinking on carbon cycling (Tracing Carbon). *Interactive Learning Environments*.
- [28] Baez, R. Sanchez, H., & Pllana, D. (2025). Fostering rigor through spiraled mathematics education. *Inter [26] Journal of Research in Education and Science (IJRES)*, 11(4), 922–943.
- [29] Klemen, T., Novak, L., & Pahor, A. (2025). Introduction of hydrosphere environmental problems in chemistry lessons through an online workshop. *Education Sciences*, 15(1), 57.
- [30] Terrell, C., & Miller, L. (2025). No more headless chickens: Using lab closure as an intervention to improve students' retention, confidence, and comprehension of laboratory concepts. *Journal of Biological Chemistry*.
- [31] Hutagalung, W. V., Ashari, E., & Sinaga, J. B. (2025). Inferential question on teaching and learning process. *Celtic: Journal of Culture, English Language Teaching, Literature & Linguistics*.
- [32] Lichty, E., Klehm, W. D., & Metellus, M. L. (2025). Procedural learning vs. conceptual mastery in undergraduate labs. *ASEE Proceedings*.
- [33] Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders*, 32(1), 7–18.
- [34] Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. Routledge.
- [35] English, L. D. (2016). STEM education K–12: Perspectives on integration. *Inter [26] Journal of STEM Education*, 3(1), 3.
- [36] Kartika, L. D. (2025). The use of visual-based teaching media in fiqh learning on the topics of fasting, Friday prayer, and sunnah prayers for secondary school students. *Giyat: Education Science*.
- [37] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

