encyclopedia

MDPI

Review

From Obstacle to Catalyst: A Narrative Review of Learning from Errors in Science Education

Noëlle Fabre-Mitjans [1,*], Gregorio Jiménez-Valverde [1,2], Salvador Viciana [1] and Ferran Crespo i Torres [1]

[1] EduCiTS Innovation and EMA Research Groups, Faculty of Education, Universitat de Barcelona, 08035 Barcelona, Spain; gregojimenez@ub.edu (G.J.-V.); salvador.viciana@ub.edu (S.V.); f.crespoitorres@ub.edu (F.C.i.T.)
[2] Institut de Recerca en Educació (IREUB), Universitat de Barcelona, 08035 Barcelona, Spain
* Correspondence: nfabre@ub.edu

Abstract

Learning from errors (LFE) constitutes a complex, multilevel process whose pedagogical value depends on how errors are elicited, interpreted, and corrected. This narrative review critically synthesizes research on LFE and its implications for science instruction in school and higher education. The reporting of the purposive, non-exhaustive search was strengthened with reference to the Scale for the Assessment of Narrative Review Articles (SANRA) criteria. The synthesis traces the historical transition from error avoidance to conditional error-as-resource perspectives and integrates five complementary lenses: error type, negative knowledge, cognitive load and prior knowledge, affective and self-regulatory processes, and the Generating Errors/Detecting and Correcting Errors model. It then evaluates instructional approaches including productive failure, erroneous examples, peer instruction, formative feedback, conceptual-change strategies, Predict–Observe–Explain, refutation texts, and gamification. Evidence is strongest when errors are embedded in structured cycles of manageable challenge, explanation, feedback, and revision; however, effects are heterogeneous and moderated by learner expertise, task demands, feedback design, classroom error climate, and safety constraints. Domain-specific evidence is most developed in physics, chemistry, biology, and health-professions simulation, whereas Earth and climate sciences and teacher professional development remain comparatively underexamined. The review concludes with an integrative framework and actionable design principles for science teacher education, while identifying methodological, cultural, and empirical gaps for future research.

Keywords: science education; learning from errors; productive failure; error climate; metacognition; feedback; peer instruction; conceptual change; teacher education

Academic Editors: Abdeljalil Métioui and Konstantinos Ravanis

Received: 27 June 2026
Revised: 26 August 2026
Accepted: 31 August 2026
Published: 2 September 2026

1. Introduction

The contemporary educational landscape is increasingly characterized by growing concerns about students' disengagement from scientific disciplines [1]. The persistence of traditional methodologies has been associated with negative perceptions regarding science [2], as well as inefficient content acquisition [3]. Furthermore, frustration constitutes a prevalent emotional response when students encounter complex scientific material [4], a phenomenon frequently associated with making errors during the learning process [5]. School science can sometimes present knowledge acquisition as a linear progression toward a single correct answer, whereas scientific practice advances through conjecture, evidential criticism, revision, and the disciplined correction of error [6,7]. Consequently, there is a need for innovative approaches and for a shift in attitudes toward the role of errors in the classroom.

Although errors have historically carried negative connotations due to their adverse impact on summative assessments [8], the analysis of errors offers significant pedagogical value. Specifically, it provides educators with critical insights into the underlying sources of student misconceptions while simultaneously enabling students to differentiate between productive and unproductive lines of reasoning [9]. Rather than conventional practices that seek to avoid errors entirely [10], evidence suggests that integrating errors into the instructional process yields educational benefits [11]. This approach has been investigated across diverse disciplines, including mathematics, medical education, and science learning [12–15]. Despite growing interest in Learning from Errors (LFE), the literature remains dispersed across educational domains and has not, to our knowledge, been integrated into a science-education synthesis that connects cognitive, affective, metacognitive, and instructional mechanisms [5,16].

In response to this literature gap, this review aims to synthesize the most relevant empirical and theoretical findings intersecting LFE with science education. Specifically, it addresses four overarching research questions: (1) How has the educational significance of errors evolved historically? (Sections 3.1–3.3), (2) Through which cognitive, emotional, and metacognitive processes can errors foster learning? (Section 4), (3) What are the primary challenges and boundary conditions associated with using errors in instruction? (Section 5), (4) Which instructional approaches leverage errors as learning opportunities, and under what conditions? (Sections 6 and 7). The review also examines practical applications across science disciplines and concludes by consolidating design principles for teacher education, highlighting systemic limitations, and identifying evidence-based priorities for future research and instructional design.

2. Scope and Approach of the Review

This article is a narrative review designed to build a conceptually coherent synthesis across educational psychology, science education, mathematics education, and adjacent high-stakes training domains. The search was purposive and non-exhaustive: sources were selected for their capacity to clarify concepts, mechanisms, boundary conditions, instructional approaches, or science-specific applications rather than to estimate a single pooled intervention effect. The revised reporting was checked against the six SANRA domains—importance, aims, search description, referencing, presentation of evidence levels, and reporting of relevant outcome data—while recognizing that SANRA is an appraisal aid rather than a substitute for a systematic-review protocol [17].

Eligibility was applied at the level of conceptual or empirical contribution (Table 1). The review included peer-reviewed journal articles, scholarly books and chapters, and selected conference proceedings published between 1950 and 2026. The final literature search and retrieval process was concluded on 5 June 2026. English-language work was prioritized, with selected non-English sources retained when historically or conceptually important. Sources had to address errors, mistakes, failure, misconceptions, corrective feedback, or closely related instructional processes in school, higher education, teacher education, or adjacent scientific training contexts. Studies concerned only with error rates, without an educational or learning interpretation, were excluded.

https://doi.org/10.3390/encyclopedia6090192

Table 1. Eligibility framework used in the narrative review.

Dimension	Included	Excluded
Publication status and period	Peer-reviewed articles; scholarly books/chapters; selected conference proceedings; 1950–2026.	Unpublished manuscripts and non-scholarly commentary, except when a historically necessary definition was unavailable elsewhere.
Educational setting	School, higher education, teacher education, and adjacent scientific or safety-critical training.	Work with no plausible instructional, learning, assessment, or training relevance.
Substantive focus	Error generation, error detection/correction, error climate, misconceptions, failure, feedback, productive failure, incorrect examples, conceptual change, or related mechanisms.	Descriptive error counts with no analysis of learning, reasoning, feedback, or instructional use.
Evidence role	Theoretical frameworks, primary empirical research, evidence syntheses, and methodologically informative historical sources.	Sources that did not materially contribute to one of the four research questions.
Language	Primarily English; selected non-English sources when historically or conceptually relevant.	Non-English sources that could not be evaluated reliably by the authors.

Literature retrieval was iterative and purposive across the Web of Science Core Collection, Scopus, and ERIC. Google Scholar was used only for supplementary backward and forward citation tracking. Searches were conducted in titles, abstracts, and keywords where database interfaces permitted. Rather than relying on a single rigid Boolean string, the search used clusters of terms that were combined according to the conceptual question being examined. Table 2 documents those clusters and the search fields used.

Table 2. Search-term clusters and their analytical purposes.

Analytical Purpose	Representative Terms
Core construct	learning from errors; learning from error; error-based learning; errorful learning; error management training; mistake failure
Cognitive and metacognitive mechanisms	negative knowledge; error generation; error detection; error correction; metacognition; self-regulation; cognitive load; prior knowledge; desirable difficulties
Instructional approaches	productive failure; problem solving before instruction; incorrect/erroneous examples; corrective/formative feedback; peer instruction; refutation text; Predict–Observe–Explain; gamification
Science and adjacent domains	science education; physics; chemistry; biology; ecology; geoscience; climate education; Nature of Science; medical/nursing simulation; forensic training

Potential sources were assessed through titles, abstracts, and, when relevant, full texts for their capacity to contribute to the conceptual framework or the domain-specific synthesis. Backward and forward citation tracking was used to identify foundational texts and newer extensions. The initial search was conducted by one of the authors and yielded 223 records. After removing sources that did not align with the focus of the review or failed to meet the inclusion criteria, the final corpus comprised 139 sources. This final screening was performed independently by two authors, who subsequently reconciled any discrepancies through discussion with the rest of the authors to achieve full consensus.

For descriptive transparency, each cited source was assigned one primary domain and one primary contribution type according to its predominant role in this review. These categories are interpretive rather than bibliometric, because several sources span more than one discipline or method. The resulting profile is shown in Table 3.

Table 3. Descriptive profile of the final cited corpus.

Classification	Primary Category	Number of Sources	Representative Contribution
Source domain	Science education and science teacher education	47	Misconceptions, conceptual change, disciplinary interventions, Nature of Science, and teacher knowledge.
Source domain	Educational psychology, learning sciences, and general education	57	Errorful generation, affect, metacognition, feedback, self-regulation, and instructional design.
Source domain	Mathematics education	22	Error climate, teacher error handling, incorrect examples, and problem-solving-before-instruction studies.
Source domain	Health, medical, forensic, and safety-critical training	6	Simulation, clinical reasoning, productive failure, and safe error exposure.
Source domain	Organizational, human-factors, and workplace learning	6	Error taxonomies, error management training, transfer, and negative consequences.
Source domain	Review methodology/reporting	1	Narrative-review quality and transparency.
Contribution type	Primary empirical research	84	Experimental, quasi-experimental, observational, survey, qualitative, and mixed-methods studies.
Contribution type	Evidence synthesis	25	Systematic/narrative reviews and meta-analyses.
Contribution type	Theoretical, conceptual, or methodological framework	14	Models of LFE, conceptual change, feedback, metacognition, and error management.
Contribution type	Books, chapters, proceedings, and reference resources	16	Foundational monographs, chapters, conference studies, and definitions.

Mathematics education supplies a substantial part of the evidence on error climate, teacher response, erroneous examples, and problem solving before instruction. Transfer to science education is therefore treated as an analytical proposition rather than an assumption. Mathematical and scientific errors can both reveal reasoning and support comparison, but scientific claims are additionally constrained by empirical evidence, representational models, material apparatus, and—in laboratory or health contexts—physical safety. Accordingly, mathematics-based findings are used to formulate candidate mechanisms and design conditions, while science-specific evidence is prioritized when disciplinary claims are made.

The review uses a qualitative evidence-strength appraisal to avoid implying a level of precision that the heterogeneous corpus cannot support. "Strong" indicates convergent evidence from meta-analytic synthesis and/or replicated findings across settings with broadly consistent conclusions; "moderate" indicates support from several empirical studies or from an evidence synthesis affected by substantial heterogeneity, inconsistency, or transfer limitations; and "emerging" indicates a small, domain-specific, indirect, or primarily conceptual evidence base. These labels are interpretive summaries rather than formal

GRADE assessments or study-level risk-of-bias assessments. The labels characterize the evidential basis for the review's synthesized claims rather than the methodological quality of individual studies.

3. Error Definitions and Historical Evolution: From Error Avoidance to Error as a Resource

3.1. Terminology

A working definition is needed before examining the educational role of error. Although a broad cross-cultural taxonomy is provided by Vanderheiden and Mayer [18], pedagogical research often uses "mistake," "error," and "failure" inconsistently. James [19] distinguishes an unintentional, performance-based lapse that a learner can self-correct from a deeper error that reflects faulty knowledge or reasoning. The present review adopts a broad definition of error as the outcome of an action that deviates from an expected standard or intended goal [20,21], while Table 4 clarifies how neighboring terms are used.

Table 4. Working terminology used throughout the review.

Term	Working Definition	Use in This Review
Mistake	An unintended performance lapse that the learner can often recognize and self-correct [19].	Used when the cited source specifically concerns slips or self-correctable performance failures.
Error	An action, response, representation, or conclusion that deviates from an expected standard or goal [20,21].	The umbrella construct used for LFE; an error does not necessarily imply a stable misconception.
Failure	Failure to attain a task goal or generate a canonical solution; it can contain several errors but is not synonymous with any single error [21,22].	Retained in established constructs such as Productive Failure.
Misconception/alternative conception	A noncanonical scientific explanation that may be relatively coherent or may arise from context-sensitive knowledge elements [23,24].	Used only when evidence supports a recurring conceptual structure, not for every wrong answer.
Negative knowledge	Knowledge of what is incorrect, ineffective, or inappropriate, together with cues for avoiding that pathway in the future [25,26].	Treated as one possible learning outcome of diagnosing and correcting errors.

Rasmussen's classification [27], developed in human-factors research, distinguishes skill-based lapses during automated activity, rule-based errors caused by misapplication or forgetting of a rule, and knowledge-based errors in unfamiliar or complex situations. Rule-based and knowledge-based errors are especially relevant to conceptual science learning, although skill-based errors remain important in laboratory procedures. Representational errors are treated in this review as a cross-cutting pedagogical form—an incorrect diagram, model, equation, or worked solution—rather than as a fourth level in Rasmussen's taxonomy.

3.2. The Legacy of Error Avoidance

Historically, behaviorist paradigms viewed errors as adverse events that should be minimized, presuming that the rehearsal of incorrect responses reinforces undesirable stimulus-response associations [10,13]. This perspective is exemplified by Skinner [28], who argued that instructing students accurately from the outset fosters effective knowledge and skill acquisition by minimizing errors through incremental, progressive steps. Within this theoretical framework, errors should be prevented because they are presumed to solidify incorrect cognitive pathways, thereby increasing the difficulty of subsequent remedial instruction [10]. Consequently, behaviorist models imply that errors can be minimized through effective listening, flawless instructional delivery, and an optimized learning environment. By suggesting that errors can be largely avoided under ideal instructional conditions, this paradigm tends to overlook the role of individual differences and the pedagogical significance of errors.

When errors occurred, behaviorist traditions used behavioral modification techniques to weaken the undesired response. These included extinction through the withdrawal of reinforcement, suppression through aversive consequences, and vicarious learning from the observed consequences of another person's response [29]. Such techniques differ substantially from contemporary approaches that ask learners to explain why an answer is wrong and how it should be revised.

Risk-averse approaches are justified when physical safety or immediate operational accuracy is non-negotiable. In ordinary science learning, however, strict error avoidance can conceal learner reasoning, restrict hypothesis generation, and remove opportunities to compare intuitive explanations with accepted scientific models [6,23]. Preventing the public expression of an error does not necessarily eliminate the underlying conception; it may simply make that conception less visible to the teacher. Conversely, constructivist and conceptual-change frameworks reframed errors as evidence of learners' attempts to make sense of experience [23]. Within these traditions, incorrect responses are valued not because error commission is intrinsically beneficial, but because it reveals how learners actively organize and structure their experiences. This shift was especially important for science education. Decades of research show that students routinely interpret scientific phenomena through their own logical, even if scientifically incorrect, frameworks [22,30].

Stevenson and Stigler's cross-national account [31] described broad differences between classroom practices in the United States and East Asia at the time, including greater public examination of student-generated solution methods in Japanese lessons. Students in the East Asian systems discussed in that work also performed better on standardized mathematics assessments. This comparison is historically suggestive, but it does not isolate error handling as the causal factor: curriculum, teacher preparation, instructional time, assessment, and cultural norms also differed.

Conceptions of science and scientific error have likewise changed over time. Science was often represented publicly as an authoritative body of finished knowledge, while errors were treated as embarrassing blunders caused by individual incompetence. Historical and philosophical analysis instead shows that error detection, criticism, and revision are recurrent features of scientific practice [6,32].

3.3. Errors as Learning Opportunities

Contemporary research does not support the view that errors are intrinsically beneficial. Benefits are most likely when the task creates a manageable rather than overwhelming difficulty, the learner activates relevant prior knowledge, and the error is followed by informative correction and an opportunity to revise [10,22,33–35]. Under these conditions, engagement with errors can support attention [36], argumentation [37], self-regulation [38], retention and critical thinking [39], and effort regulation [40]. Classroom studies also report associations between constructive error handling, achievement, and motivation [41,42], although the size and direction of effects depend on context. Paradoxically, learners may benefit from generating errors even when they do not recognize that benefit. In a study by Huelser and Metcalfe [43], students subjectively perceived that studying directly paired cue-target words was superior for their learning compared to a condition requiring initial error generation prior to receiving the correct response. However, objective performance metrics revealed that the error-generation condition yielded significantly higher retention and memorization scores.

This conceptualization of errors as instructional resources necessitates careful consideration of the broader social context. Students do not encounter errors in isolation; rather, they experience them within classroom environments where instructors and peers actively frame the meaning of those errors. Research focusing on the perceived "error climate" demonstrates that learners clearly differentiate between classrooms where errors are tolerated, analyzed, and utilized as learning opportunities, and those where errors are treated as threats to personal competence or social reputation [40]. This environmental distinction is critical, as identical cognitive discrepancies can prompt deep reflection in a supportive climate but trigger behavioral withdrawal in a punitive one.

Longitudinal and intervention studies further indicate that teacher responses shape whether learners engage with or avoid their errors. A constructive error climate can buffer declines in adaptive responses over time [44,45], and short professional-development interventions can improve students' perceptions of teacher support during errors, even when deeper anxiety and learning orientations are harder to change [46]. These findings are relevant to science education, where complex reasoning is often public and evaluative, but they should be transferred cautiously because much of the underlying evidence comes from mathematics classrooms. Scientific progress itself frequently depends on identifying and correcting flawed claims or methods [6].

4. Why Errors Can Support Learning: Cognitive, Emotional, and Self-Regulatory Mechanisms

Errors can arise from interacting sources, including gaps or conflicts in prior knowledge, excessive cognitive load, ambiguous task interpretation, limited motivation, and constraints in instructional materials or support [10,35]. Reflection on an error can contribute to "negative knowledge": knowledge of what a concept or action is not, which pathways are ineffective, and when they should be avoided [25,47]. The value lies not in preserving the error but in using the contrast between an unsuccessful pathway and a warranted alternative.

Negative knowledge can help learners anticipate the consequences of actions and connect new experiences with prior cases. Errors perceived as relevant can become memorable episodes that guide more adaptive responses in analogous situations [48]. This account is especially plausible when the learner can explain why the initial pathway failed and can retrieve that explanation later.

In a biology intervention, students who diagnosed an incorrect energy-flow representation gained more conceptual knowledge than peers who evaluated a correct representation [49]. The present review interprets this finding as compatible with selective rejection of inaccurate model components and integration of better-supported information. However, the study measured conceptual outcomes rather than internal representational "tagging"; the proposed updating mechanism should therefore be understood as an explanatory interpretation, not as a directly observed cognitive process. Oser and Spychiger [26] similarly argue that negative knowledge is embedded in broader knowledge structures and helps learners recognize why previously accepted premises are untenable.

Error management training (EMT) provides evidence that actively exploring and learning from errors can improve later adaptation. In Keith and Frese's experimental work, emotion control and metacognitive activity mediated transfer benefits [50]. A subsequent meta-analysis of 24 studies found a positive overall effect of EMT ($d = 0.44$), with larger effects for adaptive transfer to structurally different tasks ($d = 0.80$) [51]. Because most EMT studies concern workplace or computer-skill training, transfer to school science should be treated as mechanistic support rather than direct classroom evidence. Beyond purely cognitive dimensions, affective factors also shape LFE. According to the learning model formulated by Tulis et al. [5], a comprehensive instructional approach to errors must account for emotional, self-regulatory, and motivational dynamics. From a neuroscientific perspective, individual variations in error processing can serve as predictors for the risk of developing anxiety disorders [52]. This raises an important developmental question: can young children learn from failure when it is appropriately supported, or do educators sometimes limit such opportunities by over-protecting them from experiences perceived as harmful [53]? While certain researchers have noted that youth frequently conceptualize failure as inherently detrimental—associating it with negative affective states and a diminished sense of academic belongingness (e.g., [54])—others have generated contradictory evidence indicating more resilient orientations [55].

In this affective landscape, Dweck and Reppucci [56] examined "learned helplessness" as a psychological state wherein a student perceives a stressful situation as entirely beyond their locus of control. In these scenarios, students seeking to bypass the frustration of failure often solicit immediate pedagogical intervention from the instructor without attempting autonomous problem-solving; conversely, educators frequently intervene prematurely merely to mitigate student distress [53]. These situations are common in science-making contexts [57]. Furthermore, Crowe and Higgins [58] demonstrated that an individual's regulatory focus modulates how they respond to errors. Whereas individuals operating under a promotion focus are primarily oriented toward achievement and accomplishment, those with a prevention focus prioritize security, accuracy, and institutional responsibility. While these regulatory orientations often reflect stable personality traits, instructional design can manipulate them by deliberately framing how error consequences are presented in the learning environment.

A third mechanism is metacognition, understood as knowledge about cognition together with the monitoring and control of one's cognitive activity [59]. To learn from an error, students must notice a discrepancy, judge its likely cause, select a corrective strategy, and evaluate whether the revision resolves the problem. Higher education research increasingly treats strategic knowledge for learning from errors as a measurable competence, showing that learners differ not only in error frequency but also in whether they know how to respond productively after an error [60].

Corrective information alone may therefore be insufficient. Learners need structured opportunities to compare their work with criteria or exemplars, generate self-feedback, and act on that comparison through revision [61–63]. Nicol and Rose [61] found that exemplars examined after students had produced work generated more actionable and context-specific self-feedback than exemplars provided beforehand, because learners could relate the comparison to difficulties in their own work. Zhang and Fiorella's model [22] organizes learning from self-generated errors into two major phases: Generating Errors (GE), which can activate and differentiate relevant prior knowledge, and Detecting and Correcting Errors (DCE), in which comparison, self-explanation, and feedback support revision. The model also recognizes that emotion, motivation, individual differences, and design features condition both phases.

An Integrated Conceptual Architecture

The theoretical resources reviewed above are complementary rather than interchangeable. Rasmussen's taxonomy identifies the form and origin of the error; negative knowledge describes one possible learning product; cognitive-load and prior-knowledge research specify when an error can be processed; the affective and self-regulatory model explains engagement or withdrawal; and the GE/DCE model specifies a temporal sequence through which generation, detection, explanation, feedback, and correction can become productive [5,22,25,27,35]. Table 5 integrates these dimensions and carries them forward into the appraisal of instructional approaches and science domains.

https://doi.org/10.3390/encyclopedia6090192

Table 5. Integrative matrix linking error forms, mechanisms, instructional approaches, and boundary conditions.

Error Form	Manifestation	GE/DCE Mechanism	Potential Instructional Approaches	Key Moderators	Evidence Appraisal
Skill-based	Slip or lapse during a familiar procedure; e.g., incorrect instrument setting or omitted laboratory step.	DCE dominates: rapid detection, procedural feedback, rehearsal, and automation.	Immediate corrective feedback; checklists; simulation; worked examples; deliberate practice.	Physical safety, time pressure, fatigue, and whether the procedure can be practiced without real-world harm [27,64].	Moderate in training and safety-critical domains; limited direct science-classroom synthesis.
Rule-based	A known rule is forgotten, overgeneralized, or applied in the wrong context.	GE exposes the selected rule; DCE requires comparing conditions of applicability and explaining the mismatch.	Incorrect examples; peer instruction; targeted feedback; POE; refutation tasks.	Prior knowledge, clarity of rule boundaries, feedback specificity, and cognitive load [65–67].	Moderate; findings vary with prompting and learner expertise.
Knowledge-based	Reasoning in an unfamiliar or conceptually complex situation produces a noncanonical explanation or unsuccessful solution.	GE activates and differentiates prior knowledge; DCE supports conceptual reorganization, negative knowledge, and self-explanation.	Productive failure; inquiry/POE; peer argumentation; conceptual-change tasks; refutation texts.	Task complexity, age, domain knowledge, affect, scaffolding, consolidation quality, and error climate [22,35,68,69].	Moderate overall, with strong evidence for some conceptual-change strategies and substantial heterogeneity.
Representational (cross-cutting)	Incorrect diagram, model, equation, graph, or worked solution.	DCE requires locating the discrepancy, explaining why it matters, and revising the representation; negative knowledge can support future avoidance.	Erroneous examples; model critique; concept cartoons; peer instruction; visual comparison.	Risk of retaining incorrect visual information, representational competence, confidence, and explanation prompts [49,70–73].	Emerging-to-moderate; average effects of erroneous examples are small unless explanation is well designed.

Note: The matrix is an interpretive synthesis, not a claim that each instructional method maps exclusively to one error type.

https://doi.org/10.3390/encyclopedia6090192

5. Challenges and Boundary Conditions

5.1. Not All Errors Are Equally Useful

The pedagogical value of errors is conditional. Some are slips, inattentive omissions, or misreadings; others expose rule misapplication or deeper conceptual commitments. Some occur in tasks that learners can meaningfully analyze, whereas others exceed available knowledge and working-memory resources. Mera et al.'s review [35] identifies error type, prior knowledge, feedback timing and specificity, semantic relatedness, learner motivation, and outcome measure as important moderators. Thus, one synthesis should not be mistaken for evidence that all errors or all forms of struggle are educationally equivalent. Evidence from an adjacent high-stakes science domain illustrates this point. In medical education, learning from erroneous clinical case examples substantially improved reasoning, but additional prompts intended to support self-explanation increased extraneous cognitive load and were particularly problematic for learners with low self-efficacy [74]. The lesson is not that prompting is ineffective, but that error-based designs can become counterproductive when they overburden novice learners. The instructional challenge is to make errors visible enough to be educative without making the task so demanding that learners cannot interpret them.

5.2. Teacher Beliefs, Ambivalence, and Preparation

In alignment with the theory of negative knowledge and foundational empirical research, an error-tolerant classroom culture constitutes an important condition for students to effectively learn from their errors [75]. Teachers' attitudes toward errors can either support or hinder students' ability to learn from failure, depending on whether errors are treated as learning opportunities or as signs of deficiency [55,76]. The Learning Error and Formative Feedback (LEAFF) model posits that learners require an emotionally secure environment to explicitly articulate their cognitive errors or underlying misconceptions [11]. Consequently, instructors play a decisive role in establishing these supportive conditions, and their beliefs influence how they use errors in instruction [77]. Previous research suggests that teachers who adopt a behaviorist orientation often either overlook errors or correct them immediately, rather than using them as opportunities for reflection [78].

Conversely, educators adopting constructivist perspectives more often treat student errors as generative learning opportunities [79]. Students' achievement and motivation are associated with how teachers handle errors in class [41], but the available multilevel evidence does not by itself establish a simple causal effect of teacher attitudes. Even teachers who endorse the value of errors often use them only to a limited extent [79]. This gap may reflect the difficulty of valuing errors cognitively while managing the negative emotions they evoke in public classroom interaction [80]. Furthermore, Zamora et al. [81] found that early childhood and primary preservice teachers were significantly less likely than their peers in secondary teacher education to view classroom errors as learning opportunities, and reported higher levels of frustration intolerance. Crucially, these specific psychological traits were identified as robust predictors of the teachers' attitudes regarding their own professional or personal errors. These findings highlight the need for specialized teacher preparation programs designed to equip preservice teachers with effective, actionable error-management strategies for classroom implementation. Ultimately, despite the documented benefits of error engagement in fostering student self-regulation strategies [50], the systematic integration of academic error analysis within contemporary teacher training curricula has historically remained limited [82].

The cultural scope of error-climate evidence also requires caution. Much of the construct's development and longitudinal validation comes from German-speaking or broadly Western mathematics classrooms [40,44,45,75], with a smaller body of work in other national contexts such as Tanzania [46]. Norms governing public disagreement, face-saving, teacher authority, and the social meaning of an incorrect answer vary across educational cultures. A supportive error climate should therefore be operationalized and validated locally rather than assumed to function identically across settings.

5.3. Science-Specific Tensions

Science education also involves a particular tension: some errors are harmless and conceptually valuable, whereas others may have safety implications. This is especially evident in laboratory work and in health professions education, but it also matters in school science, where inaccurate handling of equipment, chemicals, or electrical setups cannot be treated simply as another opportunity for exploratory failure. For this reason, science pedagogy often needs protected environments—paper-and-pencil representations, simulations, worked examples, or virtual labs—in which conceptual or procedural errors can surface without endangering people or materials.

The distinction between productive and harmful error is therefore central. LFE in science does not mean celebrating every incorrect action. It means designing environments in which the kinds of errors worth discussing are pedagogically accessible, emotionally manageable, and physically safe. This point becomes especially clear in adjacent high-stakes science training fields, where the instructional use of error is carefully separated from the acceptance of preventable failure in real practice [64,83]. These boundary conditions explain why LFE requires designed instructional structures rather than the mere permission to make errors.

6. Instructional Approaches That Leverage Errors

The contemporary literature does not present a single error pedagogy. Rather, it offers a family of approaches that differ in how errors are elicited, who generates them, what forms of feedback are provided, and whether revision is immediate or delayed.

6.1. Productive Failure

The sequence of instruction remains contested: instruction followed by problem solving (I–PS) can reduce initial cognitive burden, whereas problem solving followed by instruction (PS–I) can prepare learners to notice critical features and knowledge gaps [69,84]. Productive Failure (PF) is a structured form of PS–I in which learners first generate incomplete or noncanonical solutions and then participate in an explicit consolidation phase. PF should not be equated with leaving novices in minimally guided discovery for an entire unit; its proposed benefits depend on the subsequent comparison, explanation, and instruction [85–87].

PF is perhaps the most influential contemporary framework for intentionally leveraging learner error. In Kapur's original formulation, students first struggle with a complex, novel problem without being shown the canonical procedure, and only afterwards receive instruction that helps them compare and consolidate solution methods [85,86]. The key claim is not that failure itself teaches, but that unsuccessful generation activates prior knowledge, differentiates problem features, and prepares learners to learn more deeply from later explanation.

Meta-analytic evidence is favorable but conditional. Sinha and Kapur's synthesis of 53 studies reported a moderate advantage of PS–I over I–PS (Hedges' $g = 0.36$, 95% CI [0.20, 0.51]), with larger effects when implementations more closely adhered to PF design principles [88]. However, younger learners in Grades 2–5 and domain-general skills showed trends favoring I–PS, and age, duration, study design, and implementation fidelity moderated outcomes. Further studies have not always replicated a PF advantage: fifth graders

https://doi.org/10.3390/encyclopedia6090192

showed no benefit over direct instruction in one boundary-condition study [89], and observing a peer's complete failed problem-solving process equaled or exceeded learners' own problem solving in another experiment [90]. These findings indicate that unsuccessful generation is not automatically the active ingredient and that PF is best treated as a design family with explicit boundary conditions.

Within science and health-science education, the evidence base is promising but smaller and context-specific. PF has supported future learning in health-professions education [91] and was associated with gains of about five percentage points on relevant topics in one large undergraduate biology quasi-experiment, with particularly visible benefits for lower-performing students [92]. These studies support further disciplinary experimentation but do not justify assuming equal effects across ages, concepts, or outcome types.

6.2. Incorrect Examples, Self-Diagnosis, and Incorrect Representations

A second family of approaches relies on providing incorrect examples. Here the pedagogical work lies in diagnosing and explaining a visible error. In physics classrooms, Yerushalmi and Polingher asked students to analyze "incorrect answers" collected from exam responses and found that the main difficulty was not merely spotting the flawed claim, but constructing an argument that explained why it was wrong and what a correct statement should look like [93]. This emphasis on argument structure is especially relevant to science, where the replacement of a wrong answer by a correct one is insufficient unless the conceptual warrant is also made explicit.

Video-based worked examples containing errors can support more accurate procedural representations and anticipatory knowledge about errors to avoid [70]. Yet the overall evidence is modest and design-sensitive. A meta-analysis of 42 papers found a statistically significant but small average benefit of erroneous examples; self-explanation prompts and instructional explanations strengthened their effectiveness relative to designs without error explanation [73]. Additional prompting can also overload novices, as shown in medical case-vignette research [74]. Reaño's STEAM intervention further illustrates the risk of unresolved confusion when deliberate errors are left uncorrected for too long [94].

Representational errors are especially relevant in science, where diagrams, models, equations, and graphs mediate abstract concepts. Wernecke et al. [49] embedded a misconception-based error in a biological energy-flow diagram; students who successfully identified and explained the inaccuracy achieved greater conceptual gains than peers who evaluated a correct diagram. The benefit was therefore conditional on diagnostic explanation, not mere exposure to an incorrect image. Nevertheless, one must account for the potential risks of exposing learners to incorrect representations, as they may inadvertently internalize and retain flawed information. Research indicates that when textual and visual data conflict, students frequently prioritize the information presented in illustrations [71,72].

Misrepresentations in children's science books [95] and misleading scientific information online [96] can be repurposed as instructional material, but only when the task requires students to justify the error and reconstruct a warranted representation. Future studies should include delayed tests and process measures to determine whether learners retain the correction rather than the visually salient error.

6.3. Peer Instruction and Dialogic Comparison

Peer instruction deserves special attention because it originated in university physics and remains one of the clearest examples of error-driven discussion becoming routine classroom practice. In Mazur's model, students first commit to an individual answer to a conceptual question, then discuss their reasoning with peers before revoting and receiving explanation [97,98]. The pedagogical power of the method lies precisely in the public visibility of competing conceptions. Students confront errors not as isolated private failures but as discussable alternatives that must be justified, criticized, and revised. An example of combining concept cartoons with peer instruction in science education is detailed by Jiménez-Valverde [99].

For science education more broadly, peer instruction matters because it links conceptual conflict with explanatory accountability. The method does not rely on humiliation or teacher correction alone. Instead, it creates a structured space in which wrong answers prompt argumentation. In disciplines where everyday intuitions are strong and persistent, this social comparison of explanatory resources can help students see why some answers feel plausible yet remain scientifically inadequate. This makes peer instruction especially relevant to science learning, where argumentation is not merely a communication skill but a core epistemic practice: students must evaluate claims, justify evidence, and compare competing explanations.

6.4. Feedback

Feedback is central to LFE, but it is not uniformly beneficial. Feedback Intervention Theory proposes that effects depend partly on where feedback directs attention; a major meta-analysis found a positive average effect but also reported that more than one third of feedback effects were negative [100]. In error-based learning, effective feedback should reduce the gap between the learner's current response and a warranted standard without shifting attention toward threat, shame, or an undifferentiated judgment of the self.

Hattie and Timperley's model distinguishes feedback at the task, process, self-regulation, and self-levels and organizes it around three questions: Where am I going? How am I going? and Where to next? [62]. For LFE, task-level information identifies what is incorrect; process-level feedback explains why the reasoning or strategy failed; and self-regulation feedback supports monitoring, help seeking, and revision. Person-focused praise or blame provides little information for correction and can redirect attention away from the task. Formative feedback is most useful when it is specific, non-evaluative, appropriately timed, and usable in a subsequent action [63].

The magnitude of feedback effects is heterogeneous. A meta-analysis of 435 studies reported an average effect of d = 0.48, but effects varied substantially with the information conveyed and were generally stronger for cognitive and motor outcomes than for motivational or behavioral outcomes [101]. Therefore, the question is not whether to provide "more feedback," but which information should be provided, at what level, and whether the learner has an opportunity and sufficient knowledge to use it.

Error-specific research further suggests that self-generated responses can create a productive basis for correction [102], particularly when feedback revisits the reasoning that produced the error and supplies progressively informative hints [103]. Evidence on timing is mixed: delayed feedback can support retention in some tasks [104], whereas other work attributes apparent delay benefits to proximity to the final test [105] or shows that timing interacts with task structure [106]. The hypercorrection effect adds another qualification: corrections can be especially memorable when a learner was highly confident in an incorrect answer [107], although high confidence can also make public correction emotionally threatening.

6.5. Other Conceptual-Change and Complementary Methodologies

In addition to the main methods discussed above, teachers use several other strategies to handle student errors. These techniques include pattern establishment, the formulation of cognitive paradoxes, cognitive modeling, conceptual association, and the deployment of analogies [29].

Conceptual-change research provides a major science-education framework for interpreting persistent incorrect ideas. The classical account treats some misconceptions as relatively coherent alternative frameworks that must become dissatisfied, intelligible, plausible, and fruitful before replacement [23]. Knowledge-in-pieces accounts instead describe reasoning as the context-sensitive activation and coordination of smaller knowledge elements. The two perspectives imply different error responses: replacing an integrated framework in the first case, or refining and reorganizing productive resources in the second [24]. Accordingly, the present review avoids assuming that every wrong answer reveals a unitary intuitive theory.

A meta-analysis of 218 primary studies found a large mean effect of science conceptual-change strategies (Hedges' $g = 1.10$), including cognitive conflict, cognitive bridging, and ontological-category approaches [108]. The wide prediction interval and substantial heterogeneity nevertheless show that strategy labels alone do not guarantee success. Instruction must diagnose the learner's starting conception, make relevant discrepancies interpretable, and provide resources for constructing a better explanation.

Refutation texts explicitly state a common misconception, reject it, and explain the accepted scientific account. Reviews and meta-analytic evidence indicate that such texts can outperform non-refutation texts in confronting scientific misconceptions [109–111], partly because the misconception and corrective explanation are co-activated [112]. Effects have been demonstrated for energy concepts [113]. Predict–Observe–Explain (POE) similarly elicits a prediction, contrasts it with an observation, and requires explanation, making prior models visible and creating an opportunity for revision [114,115]. Lastly, gamification may complement LFE methodologies when game elements are deliberately aligned with formative feedback, visible progress, low-stakes attempts, and opportunities for revision. Its pedagogical value does not lie in the mere addition of points, badges, or rewards, but in its capacity to organize feedback-rich cycles in which learners can test ideas, identify and tolerate errors, regulate their performance, revise their reasoning, and try again while maintaining engagement. This interpretation is consistent with work identifying feedback, progressiveness, experimentation, and error tolerance as key principles for gamified learning design [116,117]. Recent gamification studies in science teacher education further suggest that experience and health points systems, progress indicators, collaborative challenges, and iterative tasks can make attempts, feedback, consequences, and progress visible; however, they also show that gamification must be carefully calibrated, since reward or penalty-based mechanics can generate frustration, pressure, or disengagement when poorly aligned with learners' motivational needs (e.g., [118]).

7. Teaching Science Through LFE

The evidence is uneven across science domains. Table 6 summarizes representative designs, what they currently support, and the most visible gaps. The table distinguishes direct science-education evidence from mechanisms imported from adjacent fields.

Table 6. Representative empirical evidence and research gaps across science domains.

Domain	Representative Designs and Evidence	Current Implication	Principal Gaps
Physics and chemistry	Diagnostic studies, peer instruction, incorrect-answer analysis, laboratory experience sampling, and instructor error-framing studies [93,98,119–121].	Incorrect answers can reveal conceptual resources and support argumentation; framing should focus on ideas rather than public deficiency.	Few studies isolate the effects of error framing or compare correction sequences with delayed conceptual and affective outcomes.
Biology and ecology	Incorrect-representation experiments, writing-to-learn with peer review, model-evaluation studies, and teacher-knowledge analyses [49,122–124].	Representations and peer revision can expose misconceptions, but success depends on explanation quality, knowledge, and confidence.	Limited replication across biological topics, age groups, and long-term transfer; risk of propagating misconceptions through weak peer feedback.
Health and medical sciences	Quasi-experimental simulation studies, erroneous clinical cases, and conceptual models for safe failure [74,91,125,126].	Protected environments can preserve the epistemic value of errors while controlling real-world risk.	Transfer to ordinary science classrooms is uncertain; studies rarely compare emotional, conceptual, and safety outcomes together.
Earth, environmental, and climate sciences	Diagnostic drawing/conception studies, social-representation research, and refutation-text evidence applied to misinformation [111,127,128].	Spatial models and public claims offer rich material for diagnosis and refutation.	Very few controlled LFE interventions; limited evidence on probabilistic reasoning, systems thinking, and durable correction of climate misinformation.
Nature of Science	Historical-error cases, historical apparatus replication, storytelling, and game-based synthesis [6,129–131].	Errors can humanize science and foreground revision, evidence, and argumentation when discussion is explicit and reflective.	Need comparisons of narrative formats, measures of epistemic understanding, and evidence beyond short-term attitudes.
Science teacher education	Teacher misconception knowledge, error framing, preservice attitudes, and error-focused pedagogical content knowledge [81,121,124,132].	Teachers need diagnostic, discursive, affective, and safety-related competence, not merely positive beliefs about error.	Few longitudinal studies link preparation to later classroom practice or pupil outcomes; cultural validity remains underexamined.

https://doi.org/10.3390/encyclopedia6090192

7.1. Physics and Chemistry

Physics and chemistry offer fertile ground for LFE because students often rely on intuitive explanations of force, motion, matter, heat, electricity, and energy. Some wrong answers may reflect relatively coherent alternative frameworks, while others arise from context-sensitive knowledge elements; this coherence-versus-pieces debate remains unresolved and has direct instructional implications [23,24]. Error analysis should therefore ask what reasoning resources produced the answer rather than presume that every error is an isolated fact or a complete intuitive theory. Incorrect statements and worked contradictions can be useful when students must articulate the principle violated and justify a better account [30,93]. Recent evidence from misconception research strengthens this interpretation. A systematic review of misconceptions in natural sciences found that physics is the most frequently studied domain and that recurrent difficulties are concentrated around mechanics, thermodynamics, waves and sound, and radiation/light; importantly, the review also links these misconceptions to everyday experience, teacher preparation, decontextualized instruction, formula-heavy materials, and insufficient inquiry practices [119].

For this reason, students' wrong answers in physics and chemistry should be treated as diagnostic traces rather than as merely deficient performances. Cross-disciplinary work using diagnostic tests across physics and chemistry has shown that patterns of item difficulty can help teachers identify where misconceptions are concentrated, while related work with Indonesian students and prospective science teachers confirms that misconceptions cut across scientific domains and may persist into teacher preparation [133,134].

In laboratory-based chemistry, the affective consequences of error are equally important. Momentary assessment research in undergraduate chemistry laboratories indicates that error situations interact with situational engagement: errors may increase perceived challenge, but their educational value depends on whether students can connect theory and practice and seek support without the situation being interpreted as personal incompetence [120].

The same point appears in studies of error framing in undergraduate physics and chemistry contexts. Graduate teaching assistants can be trained to frame incorrect ideas as natural and useful for learning, yet the way this is operationalized matters: students tend to respond more positively when instructors use implicit, idea-focused framing rather than explicitly marking a student's contribution as wrong in front of peers [121].

The principal research need in physics and chemistry is comparative intervention evidence that separates error elicitation, public framing, explanation, and feedback. Delayed measures are especially important because an intervention may improve immediate correction while leaving conceptual transfer, confidence, or willingness to participate unchanged.

7.2. Biology and Ecology

In biology, LFE is particularly useful when concepts are abstract, multilevel, or easily conflated across representational systems. The energy concept is a case in point. Students commonly confuse energy flow with matter cycling or assume that plants obtain energy from soil. As previously mentioned, Wernecke and colleagues demonstrated that a deliberately incorrect biological representation can make these confusions discussable and thereby improve conceptual understanding [49]. The result is important because it suggests that biology instruction does not always need more simplified representations; it sometimes needs more interpretable discrepancies.

Related problems occur in respiration, genetics, homeostasis, ecology, and other multilevel topics where diagrams can conceal the point at which causal reasoning diverges from scientific models. Reviews document incomplete or inaccurate frameworks across these domains [135], while model-evaluation studies show that students differ in their ability to detect intentionally introduced biological errors [123]. LFE tasks are therefore most defensible when they make the target discrepancy explicit enough to analyze and require learners to revise the causal model rather than merely memorize labels. A recent review of misconceptions in biology likewise shows that students' erroneous explanations frequently involve incomplete or inaccurate conceptual frameworks across genetics, ecology, physiology, and related fields. This is relevant for LFE biology teaching because the aim is not only to replace a wrong label with a correct one, but to make visible the explanatory structure that leads students to apply biological knowledge incorrectly in new contexts [135].

One productive route is to combine error elicitation with revision. In introductory biology, writing-to-learn assignments followed by peer review have been shown to surface misconceptions and, when peer comments are sufficiently directed, to support their correction during revision; however, the same study also warns that vague peer feedback may propagate or even introduce additional misconceptions [122].

Model-based reasoning provides another pathway. Undergraduate students vary substantially in their metacognitive monitoring and error detection during biology model evaluation, and accuracy depends on subject-matter knowledge and confidence [123]. Error detection should therefore be taught as part of modeling competence rather than assumed to occur automatically. At the level of teaching expertise, the importance of anticipating student error is supported by evidence from high-school life science: teachers' subject-matter knowledge predicts student learning, but teachers' knowledge of students' misconceptions is an even stronger predictor when items contain popular wrong answers. In practical terms, the teacher's ability to foresee the likely error is itself a condition for making that error instructionally useful [124].

Biology research would benefit from replications across topics and developmental levels, delayed tests of corrected representations, and studies that trace whether peer explanations repair or propagate misconceptions. Process measures could distinguish successful error diagnosis from superficial answer replacement.

7.3. Health and Medical Sciences as High-Stakes Science Learning

A particularly developed body of adjacent intervention research comes from medical, nursing, and simulation-based education. This literature is valuable because it addresses a central tension in science teaching: how to use errors for learning without turning real practice into unacceptable risk. Simulations and case vignettes provide one answer. They create protected spaces in which learners can make conceptual, diagnostic, or procedural errors and then compare those errors against best practice [74,125,126].

These findings should not be generalized uncritically to all science classrooms, but they do reinforce a key design principle: when the real environment is too risky for exploratory error, the learning environment must be engineered so that errors remain epistemically visible while their consequences remain educationally manageable. This principle also appears in adjacent forensic training research, where errors made during training are treated as opportunities to strengthen later performance and should not simply be used to discredit the learner [64].

This lesson is transferable to school and university science only if the domain-specific risks are respected. In physics and chemistry laboratories, for example, some procedural errors must be prevented for safety reasons, whereas conceptual and representational errors can be relocated into diagrams, simulations, case narratives, peer discussion, or written explanations. The key design problem is therefore not whether errors should be allowed, but which errors can be made visible, discussed, and corrected without creating physical, emotional, or epistemic harm [120,121].

Future work should test which features of simulation-based error learning—authenticity, emotional safety, debriefing, repeated attempts, or expert modeling—are transferable to school and university science. Comparisons should also distinguish conceptual errors that can be explored safely from procedural violations that must be prevented.

7.4. Earth, Environmental, and Climate Sciences

Earth and environmental sciences are high-value contexts for LFE because learners often rely on intuitive spatial, temporal, and systems models. Vasconcelos and Paz [127] found that elementary pupils and teachers shared misconceptions about island formation and the relation between islands and the seafloor, illustrating how drawings can function as diagnostic representations. An LFE task might ask learners to draw an island, compare it with a geological cross-section, identify what is missing below sea level, and revise the model in relation to formation processes. Such a task should be treated as an illustrative design derived from diagnostic evidence, not as an intervention whose effectiveness has already been established.

Climate and environmental education adds challenges of scale, probability, causality, systems interaction, and public misinformation. Preservice teachers may conflate climate and biodiversity processes or rely on socially circulating simplifications [128]. Refutation texts can juxtapose a claim such as "climate change is ordinary weather variation" with long-term records, energy-balance reasoning, and an explanation of why short-term variability does not invalidate a trend. Meta-analytic evidence supports refutation texts across scientific misconceptions [111], but controlled LFE studies specifically addressing climate reasoning remain scarce.

Research priorities in this domain include interventions on probabilistic and systems reasoning, delayed correction of misinformation, culturally diverse climate narratives, and measures that distinguish factual correction from changes in explanatory models and decision making.

7.5. Nature of Science

The Nature of Science (NoS) is a fundamental pillar of science education. Meaningful science education must acknowledge that even the most prominent historical figures made errors. When students realize that their own academic struggles mirror the hurdles faced by prominent scientists, they perceive science as a gradual, iterative journey rather than a flawless, finished product [30].

This argument aligns with contemporary accounts of scientific practice that explicitly include error as an opportunity. García-Carmona and Acevedo-Díaz argue that science education should not restrict NoS to statements about scientific knowledge, but should also address the nature of scientific practice through principles such as questioning, experimentation, modeling, cooperation, argumentation, evaluation, and the use of errors as opportunities for learning [7].

However, NoS instruction cannot rely on the mere presence of historical errors. Reviews indicate that some NoS dimensions are persistently difficult for students and teachers, especially the tentativeness of scientific knowledge, the distinction between theories and laws, the social and cultural embeddedness of science, and the rejection of a single universal scientific method. Consequently, historical errors must be accompanied by explicit and reflective discussion of why a claim was plausible, why it became untenable, and how evidential and social processes shaped its correction [136,137].

Historical narratives can support this work when they are designed as epistemic cases rather than decorative anecdotes. Empirical research on historical storytelling shows that such narratives can improve students' understanding of imagination and creativity in science, while work on the replication of historical scientific apparatus suggests that re-enacting historical experiments may improve students' attitudes toward science and provide groundwork for richer understandings of scientific practice [129,130]. Furthermore, game-based learning (GBL) might provide a useful framework for exploring NoS, demonstrating how the correction of errors is precisely what allows scientific progress to occur [131].

The remaining gap is not a shortage of historical anecdotes but a shortage of comparative evidence about how different error narratives shape epistemic understanding. Future studies should test whether learners can transfer insights about revision, evidence, and fallibility from historical cases to contemporary scientific claims.

7.6. Design Principles for Science Teacher Education

Teacher education should translate the foregoing synthesis into observable design and discourse practices. Table 7 states eight actionable principles, links each to representative evidence, and indicates the present strength of that evidence. The strength labels follow the qualitative convention described in Section 2.

Table 7. Actionable design principles for science teacher education.

Design Principle	Action in Teacher Preparation	Representative Evidence	Evidence Strength
1. Diagnose before correcting.	Train preservice teachers to distinguish slips, rule misapplications, knowledge-based reasoning, misconceptions, and representational errors before selecting a response.	[27,78,124,132]	Moderate: convergent diagnostic and teacher-knowledge evidence, but few intervention studies.
2. Separate safe error exposure from unsafe practice.	Use simulations, diagrams, written cases, or virtual laboratories for errors that would be hazardous in real settings; explicitly teach non-negotiable safety boundaries.	[64,74,125,126]	Moderate in health/simulation; emerging for ordinary science classrooms.
3. Calibrate challenge to prior knowledge.	Design GE tasks that activate relevant resources without overwhelming novices; supply preparatory examples or scaffolds when element interactivity is high.	[35,66,67,69,87]	Moderate and heterogeneous.
4. Always provide a DCE pathway.	Require detection, causal explanation, comparison with criteria or exemplars, corrective feedback, and an opportunity to revise or retry.	[22,61–63,65,68]	Strong for feedback and comparison; moderate for specific error-task formats.
5. Protect dignity while keeping reasoning visible.	Practice revoicing incorrect ideas, depersonalizing critique, eliciting alternative explanations, and avoiding public labels of incompetence.	[40,45,80,121]	Moderate; cultural generalizability requires further study.
6. Teach explanation, not answer replacement.	Use erroneous examples, peer instruction, refutation texts, and POE tasks that require why/how explanations and evidence-based warrants.	[73,93,98,109,114]	Moderate overall; explanation prompts are a consistent moderator.
7. Connect error handling to Nature of Science.	Use historical and contemporary cases to discuss why a claim was plausible, how evidence challenged it, and how scientific communities revised knowledge.	[6,7,129,130]	Moderate for NoS understanding and attitudes; limited comparative evidence.
8. Evaluate learning, affect, and transfer separately.	Assess immediate correction, delayed conceptual understanding, confidence, participation, and transfer to novel contexts rather than assuming one outcome represents all others.	[5,35,41,51,101]	Strong rationale; implementation evidence remains uneven across science domains.

Taken together, these principles define error-focused pedagogical content knowledge as the capacity to anticipate likely errors, infer their sources, create safe opportunities for explanation and revision, and evaluate whether correction changes both scientific reasoning and learners' willingness to engage. Positive beliefs about error are necessary but insufficient unless teachers can enact these practices in real interaction.

8. Conclusions and Future Directions

8.1. Answers to the Research Questions

RQ1—Historical evolution. The educational status of error has shifted from a predominantly avoidance-oriented view, in which incorrect responses were minimized or extinguished, toward a conditional resource perspective. Contemporary approaches value errors not as ends in themselves but as diagnostic and generative events that can support learning when they are made interpretable, correctable, and emotionally manageable.

RQ2—Mechanisms. Errors can foster learning through prior-knowledge activation during GE, discrepancy detection, negative knowledge, explanation, corrective feedback, metacognitive monitoring, and self-regulated revision during DCE. Affective appraisal determines whether the learner treats the discrepancy as a manageable challenge or a threat. No single mechanism explains every benefit, and the internal updating processes proposed by theory are not always directly measured.

RQ3—Boundary conditions. The usefulness of an error depends on its type, the learner's prior knowledge, cognitive load, motivation and confidence, feedback content and timing, opportunities for revision, the classroom error climate, cultural norms, and physical or epistemic safety. Findings transferred from mathematics, workplace training, or medical simulation require explicit disciplinary justification.

RQ4—Instructional approaches. The most defensible approaches organize errors within a complete learning cycle rather than merely permitting failure. Feedback and conceptual-change research provide the broadest evidence base; PF, erroneous examples, peer instruction, refutation texts, POE, simulation, and carefully designed gamification can contribute under specified conditions. Meta-analytic evidence repeatedly shows heterogeneity, making fidelity, explanation quality, learner expertise, and outcome type central to interpretation.

8.2. Limitations of the Review

This synthesis has several limitations. First, it is a purposive narrative review rather than a systematic review. Second, no formal study-level risk-of-bias assessment or quantitative synthesis across the full corpus was undertaken. Third, the literature is heterogeneous across constructs, methods, age groups, disciplines, and outcomes, and the descriptive corpus coding in Table 3 necessarily simplifies sources that span categories. Fourth, much error-climate research originates in Western and especially German-speaking mathematics contexts, so cultural transfer remains uncertain. Fifth, science-domain coverage is uneven: medical simulation and some physics, chemistry, and biology topics are better represented than Earth, climate, and longitudinal teacher-education research. Finally, adjacent-domain evidence clarifies mechanisms but cannot substitute for direct science-education trials.

8.3. Priorities for Future Research

Future research should be organized around the gaps identified by domain rather than around a generic expectation that errors will help. Priorities include: (1) comparative studies that separate error generation, explanation, feedback, and revision; (2) delayed and transfer measures alongside immediate correction; (3) process evidence from think-aloud, interaction analysis, eye tracking, or learning analytics to test proposed mechanisms; (4) longitudinal teacher-development studies linking preparation to classroom discourse and pupil outcomes; (5) cross-cultural validation of error-climate constructs; and (6) science-specific work in Earth, environmental, and climate education.

Adaptive technologies and artificial intelligence may support timely diagnosis and individualized feedback, but this direction should be evaluated rather than presumed effective. Reviews of generative-AI feedback in higher education identify opportunities for timely, personalized support alongside concerns about accuracy, transparency, and the continuing role of human judgment [138]. A systematic review of AI-driven intelligent tutoring systems in K–12 education found generally positive learning effects, but advantages were attenuated when compared with non-intelligent tutoring and the evidence was dominated by short quasi-experimental studies [139]. Future LFE research should therefore test whether adaptive systems correctly diagnose the reasoning behind an error, provide explainable and developmentally appropriate feedback, protect data and learner agency, and improve durable scientific understanding beyond conventional alternatives.

For science education to reflect scientific practice, it should neither demand flawless performance nor romanticize failure. Its aim should be the disciplined, reflective, culturally responsive, and safe navigation of error.

Author Contributions: Conceptualization, N.F.-M., G.J.-V., S.V. and F.C.i.T.; writing—original draft preparation, N.F.-M.; writing—review and editing, N.F.-M., G.J.-V., S.V. and F.C.i.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universitat de Barcelona through the Office of the Vice-Rector for Teaching Policy and the RIMDA program, grant number 2025PID-UB/021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LFE	Learning From Errors
LEAFF	Learning Error and Formative Feedback Model
I	Instruction
PS	Problem-Solving phase
PF	Productive Failure
STEAM	Science, Technology, Engineering, Arts, and Mathematics approach
POE	Predict–Observe–Explain approach
NoS	Nature of Science
GBL	Game-Based Learning
SANRA	Scale for the Assessment of Narrative Review Articles
GE	Generating Errors phase
DCE	Detecting and Correcting Errors phase
EMT	Error Management Training
ITS	Intelligent Tutoring System

References

1. Vázquez, A.; Manassero, M.A. En defensa de las actitudes y emociones en la educación científica (II): Evidencias empíricas derivadas de la investigación. *Rev. Eureka Enseñ. Divulg. Cienc.* **2007**, *4*, 417–441. [CrossRef]
2. Vedder-Weiss, D.; Fortus, D. Adolescents' declining motivation to learn science: A follow-up study. *J. Res. Sci. Teach.* **2012**, *49*, 1057–1095. [CrossRef]
3. Athuman, J.J. Comparing the effectiveness of an inquiry-based approach to that of conventional style of teaching in the development of students' science process skills. *Int. J. Environ. Sci. Educ.* **2017**, *12*, 1797–1816.
4. King, D.; Ritchie, S.M.; Sandhu, M.; Henderson, S.; Boland, B. Temporality of emotion: Antecedent and successive variants of frustration when learning chemistry. *Sci. Educ.* **2017**, *101*, 639–672. [CrossRef]

5. Tulis, M.; Steuer, G.; Dresel, M. Learning from errors: A model of individual processes. *Frontline Learn. Res.* **2016**, *4*, 12–26. [CrossRef]
6. Allchin, D. Teaching the nature of science through scientific errors. *Sci. Educ.* **2012**, *96*, 904–926. [CrossRef]
7. García-Carmona, A.; Acevedo-Díaz, J.A. The nature of scientific practice and science education. *Sci. Educ.* **2018**, *27*, 435–455. [CrossRef]
8. Soncini, A.; Matteucci, M.C.; Butera, F. Errors: Springboard for learning or tool for evaluation? Ambivalence in teachers' error-related beliefs and practices. *Soc. Psychol. Educ.* **2024**, *27*, 1455–1479. [CrossRef]
9. Ostrowska, B.E.; Biedrzycki, K.; Chrzanowski, M.M. We learn from errors and mistakes. In *Scientific Thinking in Science Education*; Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej w Krakowie: Kraków, Poland, 2021; pp. 45–54.
10. Metcalfe, J. Learning from errors. *Annu. Rev. Psychol.* **2017**, *68*, 465–489. [CrossRef] [PubMed]
11. Leighton, J.P.; Chu, M.W.; Seitz, P. Errors in student learning and assessment: The learning errors and formative feedback (LEAFF) model. In *Informing the Practice of Teaching Using Formative and Interim Assessment: A Systems Approach*; Lissitz, R.W., Ed.; Information Age Publishing: Charlotte, NC, USA, 2013; pp. 185–208. [CrossRef]
12. Hämmerle, L.; Möller, A.; Bergmann-Gering, A.; Krause-Wichmann, T.; Lederman, J. Investigating students' awareness of their own and others' deviations from controlled science experiments. *Int. J. Sci. Educ.* **2026**, *48*, 1197–1219. [CrossRef]
13. Elagha, N.; Pellegrino, J.W. Understanding error patterns in students' solutions to linear function problems to design learning interventions. *Learn. Instr.* **2024**, *92*, 101895. [CrossRef]
14. Fiori, C.; Zuccheri, L. An experimental research on error patterns in written subtraction. *Educ. Stud. Math.* **2005**, *60*, 323–331. [CrossRef]
15. Lohre, R.; Lobo, A.; Bois, A.; Pollock, J.; Lapner, P.; Athwal, G.; Goel, D. Errors in implant orientation in novice versus experienced surgeons: The positive impact of hands-on learning. *Orthop. Proc.* **2022**, *104*, 30.
16. Narciss, S.; Alemdag, E. Learning from errors and failure in educational contexts: New insights and future directions for research and practice. *Br. J. Educ. Psychol.* **2025**, *95*, 197–218. [CrossRef] [PubMed]
17. Baethge, C.; Goldbeck-Wood, S.; Mertens, S. SANRA—A scale for the quality assessment of narrative review articles. *Res. Integr. Peer Rev.* **2019**, *4*, 5. [CrossRef] [PubMed]
18. Vanderheiden, E.; Mayer, C.H. *Mistakes, Errors and Failures Across Cultures*; Springer: Cham, Switzerland, 2020. [CrossRef]
19. James, C. *Errors in Language Learning and Use: Exploring Error Analysis*; Longman: London, UK, 1998.
20. APA Dictionary of Psychology. 2019. Available online: https://dictionary.apa.org/error (accessed on 3 June 2026).
21. Simpson, A.; Maltese, A.V.; Anderson, A.; Sung, E. Failures, errors, and mistakes: A systematic review of the literature. In *Mistakes, Errors and Failures Across Cultures: Navigating Potentials*; Vanderheiden, E., Mayer, C.-H., Eds.; Springer: Cham, Switzerland, 2020; pp. 347–362.
22. Zhang, Q.; Fiorella, L. An integrated model of learning from errors. *Educ. Psychol.* **2023**, *58*, 18–34. [CrossRef]
23. Posner, G.J.; Strike, K.A.; Hewson, P.W.; Gertzog, W.A. Accommodation of a scientific conception: Toward a theory of conceptual change. *Sci. Educ.* **1982**, *66*, 211–227. [CrossRef]
24. Özdemir, G.; Clark, D.B. An overview of conceptual change theories. *Eurasia J. Math. Sci. Technol. Educ.* **2007**, *3*, 351–361. [CrossRef] [PubMed]
25. Oser, F.; Näpflin, C.; Hofer, C.; Aerni, P. Towards a theory of negative knowledge (NK): Almost-mistakes as drivers of episodic memory amplification. In *Human Fallibility: The Ambiguity of Errors for Work and Learning*; Bauer, J., Harteis, C., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 53–70. [CrossRef]
26. Oser, F.; Spychiger, M. *Lernen ist Schmerzhaft: Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur*; Beltz: Weinheim, Germany, 2005.
27. Rasmussen, J. Human errors: A taxonomy for describing human malfunction in industrial installations. *J. Occup. Accid.* **1982**, *4*, 311–333. [CrossRef]
28. Skinner, B.F. Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science* **1958**, *128*, 969–977. [CrossRef] [PubMed]
29. Türkdoğan, A.; Baki, A.; Çepni, S. The anatomy of mistakes: Categorizing students' mistakes in mathematics within learning theories. *Turk. J. Comput. Math. Educ.* **2009**, *1*, 13–26.
30. Métioui, A. Analysing the errors of renowned physicists and chemists throughout history and those of students before and after learning about science. *Encyclopedia* **2026**, *6*, 33. [CrossRef]
31. Stevenson, H.; Stigler, J.W. *The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education*; Simon & Schuster: New York, NY, USA, 1992.
32. Allchin, D. *Toward a Philosophy of Error in Science*; Oxford University Press: Oxford, UK, 2025. [CrossRef]
33. Bjork, R.A.; Bjork, E.L. Desirable difficulties in theory and practice. *J. Appl. Res. Mem. Cogn.* **2020**, *9*, 475–479. [CrossRef]
34. Kornell, N.; Hays, M.J.; Bjork, R.A. Unsuccessful retrieval attempts enhance subsequent learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **2009**, *35*, 989–998. [CrossRef] [PubMed]

35. Mera, Y.; Rodríguez, G.; Marín-García, E. Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychon. Bull. Rev.* **2022**, *29*, 753–765. [CrossRef] [PubMed]
36. Philip, A.P.; Bennett, D. Using deliberate mistakes to heighten student attention. *J. Univ. Teach. Learn. Pract.* **2021**, *18*, 193–212. [CrossRef]
37. Zhuang, Y.; Conner, A. Secondary mathematics teachers' use of students' incorrect answers in supporting collective argumentation. *Math. Think. Learn.* **2024**, *26*, 208–231. [CrossRef]
38. Zamora, Á.; Suárez, J.M.; Ardura, D. A model of the role of error detection and self-regulation in academic performance. *J. Educ. Res.* **2018**, *111*, 595–602. [CrossRef]
39. van Peppen, L.M.; Verkoeijen, P.P.J.L.; Heijltjes, A.E.G.; Janssen, E.M.; van Gog, T. Enhancing students' critical thinking skills: Is comparing correct and erroneous examples beneficial? *Instr. Sci.* **2021**, *49*, 747–777. [CrossRef]
40. Steuer, G.; Rosentritt-Brunn, G.; Dresel, M. Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemp. Educ. Psychol.* **2013**, *38*, 196–210. [CrossRef]
41. Käfer, J.; Kuger, S.; Klieme, E.; Kunter, M. The significance of dealing with mistakes for student achievement and motivation: Results of doubly latent multilevel analyses. *Eur. J. Psychol. Educ.* **2019**, *34*, 731–753. [CrossRef]
42. Tulis, M.; Steuer, G.; Dresel, M. Positive beliefs about errors as an important element of adaptive individual dealing with errors during academic learning. *Educ. Psychol.* **2018**, *38*, 139–158. [CrossRef]
43. Huelser, B.J.; Metcalfe, J. Making related errors facilitates learning, but learners do not know it. *Mem. Cogn.* **2012**, *40*, 514–527. [CrossRef] [PubMed]
44. Dresel, M.; Daumiller, M.; Spear, J.; Janke, S.; Dickhäuser, O.; Steuer, G. Learning from errors in mathematics classrooms: Development over 2 years in dependence of perceived error climate. *Br. J. Educ. Psychol.* **2025**, *95*, 180–196. [CrossRef] [PubMed]
45. Tulis, M. Error management behavior in classrooms: Teachers' responses to student mistakes. *Teach. Teach. Educ.* **2013**, *33*, 56–68. [CrossRef]
46. Kyaruzi, F.; Strijbos, J.-W.; Ufer, S. Impact of a short-term professional development teacher training on students' perceptions and use of errors in mathematics learning. *Front. Educ.* **2020**, *5*, 559122. [CrossRef]
47. von Glasersfeld, E. A constructivist's view of learning and teaching. In *Research in Physics Learning: Theoretical Issues and Empirical Studies*; Duit, R., Goldberg, F., Niedderer, H., Eds.; IPN: Kiel, Germany, 1991; pp. 29–39.
48. Kolodner, J.L. *Case-Based Reasoning*; Morgan Kaufmann Publishers: San Mateo, CA, USA, 1993.
49. Wernecke, U.; Schütte, K.; Schwanewedel, J.; Harms, U. Enhancing conceptual knowledge of energy in biology with incorrect representations. *CBE Life Sci. Educ.* **2018**, *17*, ar5. [CrossRef] [PubMed]
50. Keith, N.; Frese, M. Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *J. Appl. Psychol.* **2005**, *90*, 677–691. [CrossRef] [PubMed]
51. Keith, N.; Frese, M. Effectiveness of error management training: A meta-analysis. *J. Appl. Psychol.* **2008**, *93*, 59–69. [CrossRef] [PubMed]
52. Hajcak, G. What we've learned from mistakes. *Curr. Dir. Psychol. Sci.* **2012**, *21*, 101–106. [CrossRef]
53. Lottero-Perdue, P.S.; Parry, E.A. Perspectives on failure in the classroom by elementary teachers new to teaching engineering. *J. Pre-Coll. Eng. Educ. Res.* **2017**, *7*, 47–67. [CrossRef]
54. Heyd-Metzuyanim, E. Vicious cycles of identifying and mathematizing: A case study of the development of mathematical failure. *J. Learn. Sci.* **2015**, *24*, 504–549. [CrossRef]
55. Simpson, A.; Anderson, A.; Maltese, A.V.; Goeke, M. "I'm going to fail": How youth interpret failure across contextual boundaries. In *Proceedings of the 13th International Conference of the Learning Sciences*; Kay, J., Luckin, R., Eds.; University College London: London, UK, 2018; Volume 2, pp. 981–984.
56. Dweck, C.S.; Reppucci, N.D. Learned helplessness and reinforcement responsibility in children. *J. Pers. Soc. Psychol.* **1973**, *25*, 109–116. [CrossRef]
57. Simpson, A.; Anderson, A.; Maltese, A.V. Caught on camera: Youth and educators' noticing of and responding to failure within making contexts. *J. Sci. Educ. Technol.* **2019**, *28*, 480–492. [CrossRef]
58. Crowe, E.; Higgins, E.T. Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organ. Behav. Hum. Decis. Process.* **1997**, *69*, 117–132. [CrossRef]
59. Flavell, J.H. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *Am. Psychol.* **1979**, *34*, 906–911. [CrossRef]
60. Pfost, M.; Hübner, V. Assessment of strategic knowledge of learning from errors in higher education. *Diagnostica* **2025**, *71*, 53–63. [CrossRef]
61. Nicol, D.; Rose, J. Promoting learner self-regulation: Is it better to give students exemplars before or after producing work? *Assess. Eval. High. Educ.* **2025**, *50*, 1311–1331. [CrossRef]
62. Hattie, J.; Timperley, H. The power of feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [CrossRef]
63. Shute, V.J. Focus on formative feedback. *Rev. Educ. Res.* **2008**, *78*, 153–189. [CrossRef]

64. Eldridge, H.; Stimac, J.; Vanderkolk, J. The benefits of errors during training. *Forensic Sci. Int. Synergy* **2022**, *4*, 100207. [CrossRef] [PubMed]
65. Heemsoth, T.; Heinze, A. Secondary school students learning from reflections on the rationale behind self-made errors: A field experiment. *J. Exp. Educ.* **2016**, *84*, 98–118. [CrossRef]
66. Große, C.S.; Renkl, A. Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learn. Instr.* **2007**, *17*, 612–634. [CrossRef]
67. Durkin, K.; Rittle-Johnson, B. The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learn. Instr.* **2012**, *22*, 206–214. [CrossRef]
68. Loibl, K.; Leuders, T. How to make failure productive: Fostering learning from errors through elaboration prompts. *Learn. Instr.* **2019**, *62*, 1–10. [CrossRef]
69. Loibl, K.; Roll, I.; Rummel, N. Towards a theory of when and how problem solving followed by instruction supports learning. *Educ. Psychol. Rev.* **2017**, *29*, 693–715. [CrossRef]
70. Cattaneo, A.A.P.; Boldrini, E. You learn by your mistakes. Effective training strategies based on the analysis of video-recorded worked-out examples. *Vocat. Learn.* **2017**, *10*, 1–26. [CrossRef]
71. Peeck, J. Trends in the delayed use of information from an illustrated text. In *Knowledge Acquisition from Text and Pictures*; Mandl, H., Levin, J.R., Eds.; Elsevier: Amsterdam, The Netherlands, 1989; pp. 263–277. [CrossRef]
72. Crisp, V.; Sweiry, E. Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educ. Res.* **2006**, *48*, 139–154. [CrossRef]
73. Alemdag, E.; Eichelmann, A.; Narciss, S. A framework for learning from erroneous examples and meta-analysis of empirical research. *Rev. Educ. Res.* **2025**, *OnlineFirst*. [CrossRef]
74. Klein, M.; Otto, B.; Fischer, M.R.; Stark, R. Fostering medical students' clinical reasoning by learning from errors in clinical case vignettes: Effects and conditions of additional prompting procedures to foster self-explanations. *Adv. Health Sci. Educ.* **2019**, *24*, 331–351. [CrossRef] [PubMed]
75. Steuer, G.; Dresel, M. A constructive error climate as an element of effective learning environments. *Psychol. Test Assess. Model.* **2015**, *57*, 262–275.
76. Maltese, A.; Simpson, A.; Anderson, A. Failing to learn: The impact of failures during making activities. *Think. Ski. Creat.* **2018**, *30*, 116–124. [CrossRef]
77. Leighton, J.P.; Guo, Q.; Tang, W. Measuring preservice teachers' attitudes towards mistakes in learning environments. *Learn. Environ. Res.* **2022**, *25*, 287–304. [CrossRef]
78. Bray, W.S. A collective case study of the influence of teachers' beliefs and knowledge on error-handling practices during class discussion of mathematics. *J. Res. Math. Educ.* **2011**, *42*, 2–38. [CrossRef]
79. Pan, S.C.; Sana, F.; Samani, J.; Cooke, J.; Kim, J.A. Learning from errors: Students' and instructors' practices, attitudes, and beliefs. *Memory* **2020**, *28*, 1105–1122. [CrossRef] [PubMed]
80. Ingram, J.; Pitt, A.; Baldry, F. Handling errors as they arise in whole-class interactions. *Res. Math. Educ.* **2015**, *17*, 183–197. [CrossRef]
81. Zamora, Á.; Ardura, D.; Calderón-Lima, R. Exploring attitudes towards errors and frustration intolerance in pre-service teachers. *Teach. Teach. Educ.* **2025**, *158*, 104964. [CrossRef]
82. Rach, S.; Ufer, S.; Heinze, A. Learning from errors: Effects of teachers training on students' attitudes towards and their individual use of errors. *PNA* **2013**, *8*, 21–30. [CrossRef]
83. Frese, M.; Keith, N. Action errors, error management, and learning in organizations. *Annu. Rev. Psychol.* **2015**, *66*, 661–687. [CrossRef] [PubMed]
84. Schwartz, D.L.; Martin, T. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cogn. Instr.* **2004**, *22*, 129–184. [CrossRef] [PubMed]
85. Kapur, M. Productive failure. *Cogn. Instr.* **2008**, *26*, 379–424. [CrossRef]
86. Kapur, M.; Bielaczyc, K. Designing for productive failure. *J. Learn. Sci.* **2012**, *21*, 45–83. [CrossRef]
87. Kirschner, P.A.; Sweller, J.; Clark, R.E. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* **2006**, *41*, 75–86. [CrossRef] [PubMed]
88. Sinha, T.; Kapur, M. When problem solving followed by instruction works: Evidence for productive failure. *Rev. Educ. Res.* **2021**, *91*, 761–798. [CrossRef]
89. Mazziotti, C.; Rummel, N.; Deiglmayr, A.; Loibl, K. Probing boundary conditions of Productive Failure and analyzing the role of young students' collaboration. *npj Sci. Learn.* **2019**, *4*, 2. [CrossRef] [PubMed]
90. Hartmann, C.; van Gog, T.; Rummel, N. Preparatory effects of problem solving versus studying examples prior to instruction. *Instr. Sci.* **2021**, *49*, 1–21. [CrossRef]

91. Steenhof, N.; Woods, N.N.; van Gerven, P.W.M.; Mylopoulos, M. Productive failure as an instructional approach to promote future learning. *Adv. Health Sci. Educ.* **2019**, *24*, 739–749. [CrossRef] [PubMed]
92. Chowrira, S.G.; Smith, K.M.; Dubois, P.J.; Roll, I. DIY productive failure: Boosting performance in a large undergraduate biology course. *npj Sci. Learn.* **2019**, *4*, 1. [CrossRef] [PubMed]
93. Yerushalmi, E.; Polingher, C. Guiding students to learn from mistakes. *Phys. Educ.* **2006**, *41*, 532–538. [CrossRef]
94. Reaño, C. Error-driven learning: A new teaching methodology for improving the learning experience in STEM education. In Proceedings of the 13th International Conference on Education and New Learning Technologies (EDULEARN21), Palma de Mallorca, Spain, 5–6 July 2021; pp. 4254–4259.
95. Yilmaz, M.M.; Ozen Uyar, R.; Aslan, D. Misrepresentation of science concepts in Turkish picture books. *Issues Educ. Res.* **2020**, *30*, 1183–1203.
96. Acar, B.; Ince, E. Internet as a source of misconception. *Turk. Online J. Educ. Technol.* **2010**, *9*, 94–100.
97. Mazur, E. *Peer Instruction: A User's Manual*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
98. Crouch, C.H.; Mazur, E. Peer instruction: Ten years of experience and results. *Am. J. Phys.* **2001**, *69*, 970–977. [CrossRef]
99. Jiménez-Valverde, G. Instrucción entre iguales con viñetas conceptuales: Una propuesta en el grado de maestro de educación primaria en un contexto indagativo. In *Innovación en la Docencia e Investigación de las Ciencias Sociales y de la Educación*; Dykinson: Madrid, Spain, 2021; pp. 1511–1530. Available online: https://hdl.handle.net/2445/215654 (accessed on 15 June 2026).
100. Kluger, A.N.; DeNisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* **1996**, *119*, 254–284. [CrossRef]
101. Wisniewski, B.; Zierer, K.; Hattie, J. The power of feedback revisited: A meta-analysis of educational feedback research. *Front. Psychol.* **2020**, *10*, 3087. [CrossRef] [PubMed]
102. Grimaldi, P.J.; Karpicke, J.D. When and why do retrieval attempts enhance subsequent encoding? *Mem. Cogn.* **2012**, *40*, 505–513. [CrossRef] [PubMed]
103. Finn, B.; Metcalfe, J. Scaffolding feedback to maximize long-term error correction. *Mem. Cogn.* **2010**, *38*, 951–961. [CrossRef] [PubMed]
104. Butler, A.C.; Roediger, H.L., III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem. Cogn.* **2008**, *36*, 604–616. [CrossRef] [PubMed]
105. Metcalfe, J.; Kornell, N.; Finn, B. Delayed versus immediate feedback in children's and adults' vocabulary learning. *Mem. Cogn.* **2009**, *37*, 1077–1087. [CrossRef] [PubMed]
106. Maddox, W.T.; Ashby, F.G.; Bohil, C.J. Delayed feedback effects on rule-based and information-integration category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **2003**, *29*, 650–662. [CrossRef] [PubMed]
107. Butterfield, B.; Metcalfe, J. The correction of errors committed with high confidence. *Metacogn. Learn.* **2006**, *1*, 69–84. [CrossRef]
108. Pacaci, C.; Ustun, U.; Ozdemir, O.F. Effectiveness of conceptual change strategies in science education: A meta-analysis. *J. Res. Sci. Teach.* **2024**, *61*, 1263–1325. [CrossRef]
109. Özkan, Ö.; Tekkaya, C.; Geban, Ö. Facilitating conceptual change in students' understanding of ecological concepts. *J. Sci. Educ. Technol.* **2004**, *13*, 95–105. [CrossRef]
110. Tippett, C.D. Refutation text in science education: A review of two decades of research. *Int. J. Sci. Math. Educ.* **2010**, *8*, 951–970. [CrossRef]
111. Danielson, R.W.; Jacobson, N.G.; Patall, E.A.; Sinatra, G.M.; Adesope, O.O.; Kennedy, A.A.U.; Bhat, B.H.; Ramazan, O.; Akinrotimi, B.; Nketah, G.; et al. The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis. *Educ. Psychol.* **2025**, *60*, 23–47. [CrossRef]
112. van den Broek, P.; Kendeou, P. Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Appl. Cogn. Psychol.* **2008**, *22*, 335–351. [CrossRef]
113. Diakidoy, I.-A.N.; Kendeou, P.; Ioannides, C. Reading about energy: The effects of text structure in science learning and conceptual change. *Contemp. Educ. Psychol.* **2003**, *28*, 335–356. [CrossRef]
114. Mustofa, H.A.; Sandanadas, E.; Mohtar, L.E.; Faresta, R.A. Identifying student prior knowledge and conceptual change profiles on Newton's laws using POE (predict, observe, and explain) as a probing understanding strategy. *J. Sci. Math. Lett.* **2024**, *12*, 80–91. [CrossRef]
115. White, R.; Gunstone, R. *Probing Understanding*; Routledge: London, UK, 2014.
116. Chu, M.-W.; Fowler, T. Gamification of formative feedback in language arts and mathematics classrooms: Application of the Learning Error and Formative Feedback (LEAFF) model. *Int. J. Game-Based Learn.* **2020**, *10*, 1–18. [CrossRef]
117. Llorens-Largo, F.; Gallego-Durán, F.J.; Villagrá-Arnedo, C.J.; Compañ-Rosique, P.; Satorre-Cuerda, R.; Molina-Carmona, R. Gamification of the learning process: Lessons learned. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2016**, *11*, 227–234. [CrossRef]
118. Fabre Mitjans, N.; Jiménez-Valverde, G.; Calafell Subirà, G. ¿Gamificación digital o analógica? Comparación en cuanto a motivación y aprendizaje percibidos, ansiedad y adicción. *Profr. Rev. Currículum Form. Profr.* **2026**, *30*, 1–23. [CrossRef]

119. Guerra-Reyes, F.; Guerra-Dávila, E.; Naranjo-Toro, M.; Basantes-Andrade, A.; Guevara-Betancourt, S. Misconceptions in the learning of natural sciences: A systematic review. *Educ. Sci.* **2024**, *14*, 497. [CrossRef]
120. Kyynäräinen, R.; Vilhunen, E.; Vesterinen, V.-M. How making mistakes shapes students' situational engagement in chemistry laboratory work. *Int. J. Sci. Educ.* **2026**, *48*, 1151–1171. [CrossRef]
121. Wan, T.; Doty, C.M.; Geraets, A.; Saitta, E.K.H.; Chini, J.J. Responding to incorrect ideas: Science graduate teaching assistants' operationalization of error framing and undergraduate students' perception. *Int. J. STEM Educ.* **2023**, *10*, 5. [CrossRef]
122. Halim, A.S.; Finkenstaedt-Quinn, S.A.; Olsen, L.J.; Gere, A.R.; Shultz, G.V. Identifying and remediating student misconceptions in introductory biology via writing-to-learn assignments and peer review. *CBE Life Sci. Educ.* **2018**, *17*, ar28. [CrossRef] [PubMed]
123. Dauer, J.M.; Behrendt, M.G.; Elliott, M.; Gettings, B.; Long, T.; Clark, C. Individual variation in undergraduate student metacognitive monitoring and error detection during biology model evaluation. *Front. Educ.* **2024**, *9*, 1356626. [CrossRef]
124. Chen, C.; Sonnert, G.; Sadler, P.M.; Sunbury, S. The impact of high school life science teachers' subject matter knowledge and knowledge of student misconceptions on students' learning. *CBE Life Sci. Educ.* **2020**, *19*, ar9. [CrossRef] [PubMed]
125. Palominos, E.; Levett-Jones, T.; Power, T.; Alcorn, N.; Martinez-Maldonado, R. Measuring the impact of productive failure on nursing students' learning in healthcare simulation: A quasi-experimental study. *Nurse Educ. Today* **2021**, *101*, 104871. [CrossRef] [PubMed]
126. Palominos, E.; Levett-Jones, T.; Power, T.; Martinez-Maldonado, R. A conceptual model to inform the design of healthcare simulations that promote errors as a catalyst for learning: A discussion paper. *Nurse Educ. Pract.* **2022**, *65*, 103500. [CrossRef] [PubMed]
127. Vasconcelos, C.; Paz, M. Inquiring children and elementary school teachers to diagnose their conceptions about islands. *Front. Educ.* **2023**, *8*, 1115984. [CrossRef]
128. Calafell-Subirà, G.; Esparza, M.; Jiménez-Valverde, G. Pre-service teachers' social representations of climate change and biodiversity: Barriers and opportunities for transformative education for sustainable development. *Int. J. Sustain. High. Educ.* **2026**, *ahead of print*. [CrossRef]
129. Alisir, Z.N.; Irez, S. The effect of replicating historical scientific apparatus on high school students' attitudes towards science and their understanding of nature of science. *Sci. Educ.* **2020**, *29*, 1201–1234. [CrossRef]
130. Williams, C.T.; Rudge, D.W. Effects of historical storytelling on student understanding of nature of science. *Sci. Educ.* **2019**, *28*, 1105–1133. [CrossRef]
131. Jiménez-Valverde, G.; Fabre-Mitjans, N.; Guimerà-Ballesta, G. Games and playful activities to learn about the nature of science. *Encyclopedia* **2025**, *5*, 193. [CrossRef]
132. Degirmenci, S. Examination of prospective science teachers' pedagogical content knowledge in terms of their approaches to error: An example of content knowledge about electric circuits. *Shanlax Int. J. Educ.* **2022**, *10*, 63–87. [CrossRef]
133. Soeharto, S.; Csapó, B. Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon* **2021**, *7*, e08352. [CrossRef] [PubMed]
134. Soeharto, S.; Csapó, B. Exploring Indonesian student misconceptions in science concepts. *Heliyon* **2022**, *8*, e10720. [CrossRef] [PubMed]
135. Herak, R.; Mahanal, S.; Zubaidah, S.; Novianti, V. Exploring misconceptions in biology learning: A systematic literature review. *Salud Cienc. Tecnol.* **2025**, *5*, 1743. [CrossRef]
136. Cofré, H.; Núñez, P.; Santibáñez, D.; Pavez, J.M.; Valencia, M.; Vergara, C. A critical review of students' and teachers' understandings of nature of science. *Sci. Educ.* **2019**, *28*, 205–248. [CrossRef]
137. Khishfe, R. Improving students' conceptions of nature of science: A review of the literature. *Sci. Educ.* **2023**, *32*, 1887–1931. [CrossRef]
138. Lee, S.S.; Moore, R.L. Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learn.* **2024**, *28*, 82–106. [CrossRef]
139. Létourneau, A.; Deslandes Martineau, M.; Charland, P.; Karran, J.A.; Boasen, J.; Léger, P.-M. A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Sci. Learn.* **2025**, *10*, 29. [CrossRef] [PubMed]