Uniglobal of Journal Social Sciences and Humanities
Journal Homepage: www.ujssh.com

Design Strategies and Implementation Pathways for Interdisciplinary Science Education in Junior Secondary Schools

Huang, XiaoJuan[1*], Ramakrishnan, Kamalanathan M.[1], Dong, ZuoBao[1] & Xie, Xin[1]

[1]SEGi University, Kota Damansara, 47810 Petaling Jaya, Selangor, Malaysia

*Corresponding author email: 23259043@qq.com

To link to this article: https://doi.org/10.53797/ujssh.v5i2.25.2026

Abstract: Interdisciplinary science education requires more than placing several school subjects under a shared theme. Its central design challenge is to make disciplinary knowledge jointly necessary for explaining a phenomenon or evaluating a decision, while preserving conceptual depth and classroom feasibility. This article develops an evidence-informed framework for junior secondary science through a purposive synthesis of interdisciplinary assessment, integrated STEM education, modelling, teacher collaboration and educational design research. The framework introduces two connected design tests. Disciplinary necessity asks whether removing a disciplinary contribution changes the warranted explanation. Evidence dependency asks whether findings from one task constrain a later claim or decision. These tests are operationalised through a proposed twelve-lesson unit on school rainwater and water-quality reasoning. The unit includes an evidence-dependency map, a shared decision dossier, three organisational implementation pathways, an analytic rubric and a staged evaluation protocol. A clearly labelled synthetic teaching dataset illustrates how optical measurements, dissolved-substance reasoning, biological uncertainty and mathematical comparison can jointly change a recommendation. The design separates model-filter performance from water-safety certification and prevents missing safety evidence from being offset by a favourable cost or appearance score. No classroom implementation or student outcomes are reported. The contribution is a practical and testable specification for distinguishing substantive integration from thematic juxtaposition, together with the governance, assessment and research conditions needed before claims of educational effectiveness can be made.

Keywords: interdisciplinary science education, junior secondary schools, integrated STEM, evidence-based reasoning, curriculum design

1. Introduction

Interdisciplinary science education is often visible through its products: a model, exhibition, poster or proposed solution to a local problem. Yet the presence of several products or subject labels does not establish that students have connected disciplinary explanations. A water-themed project can contain a chemistry demonstration, a biological fact sheet and a mathematical graph while leaving the three contributions intellectually independent. The design problem is therefore how to make disciplinary knowledge interact so that it improves an explanation or changes a justified decision. Including more subjects does not solve it. A systematic review of 40 studies of interdisciplinary teaching in secondary schools found that few practices are used and fewer still achieve a real integration of disciplines. Positive effects appeared mainly where subjects were strongly linked conceptually and disciplinary and interdisciplinary moments complemented each other (Tonnetti & Lentillon-Kaestner, 2023).

The distinction is consistent with concerns about disciplinary representation and purposeful integration in STEM education (English, 2016; Kelley & Knowles, 2016). Roehrig, Dare, Ellis, and Ring-Whalen (2021) provide a detailed framework that distinguishes forms and features of integration. McLure, Tang, and Williams (2022), reviewing middle- and high-school projects, examine how integrated STEM is realised in classroom work. These sources support looking beyond an appealing context to the intellectual structure of a task. They do not imply that all interdisciplinary science must adopt an engineering-centred STEM format.

**Corresponding author: 23259043@qq.com*
https://ujssh.com/ All right reserved.

The policy context can also create an implementation imperative. China's 2022 compulsory education curriculum programme specifies that, in principle, at least 10% of each course's lesson time should support interdisciplinary thematic learning (Ministry of Education of the People's Republic of China, 2022). This article treats that historical policy as one relevant context, not as evidence of uniform implementation or as a current compliance assessment for a particular school. Chinese commentators describe the response as an attempt to integrate interdisciplinary and STEM learning within the existing subject-based curriculum rather than through separate courses (Cui et al., 2026). Design studies have also examined how integrated STEM courses can be organised around the national science, mathematics and information-technology standards for middle schools (Deng, Shen, Hao, & Zhang, 2025). A bibliometric analysis nevertheless finds that Chinese research on interdisciplinary teaching and learning has concentrated on macro-level aims, and identifies teacher development and assessment design as priorities (Chen, Yin, & Wang, 2025). The design is intended to be adaptable across junior secondary settings with different subject structures, timetables and resources.

Three questions organise the article. What distinguishes substantive disciplinary integration from shared-theme activity? How can that distinction guide a feasible sequence of tasks and teacher coordination? What assessment and research evidence would be needed to evaluate the resulting design? The article answers these questions through conceptual synthesis and a worked twelve-lesson unit. It does not report a school intervention, fabricate student responses or claim that a conceptual framework has produced learning gains.

The intended learners are junior secondary pupils who can interpret simple tables and graphs and have introductory knowledge of mixtures, particles, living organisms and fair comparisons. These prerequisites are planning assumptions to be checked locally rather than universal age-based expectations. The worked context is school rainwater and the question of why apparently clear water may still lack the evidence needed for a proposed use. Classroom work concerns models, prepared teaching materials and supplied information, not the certification or deployment of a real water-treatment system.

2. Conceptual Foundations

2.1 Integration as an explanatory relationship

In this article, interdisciplinary science means the coordinated use of concepts, methods or evidential standards from at least two disciplines to develop an explanation or judgement that is not adequately warranted by one contribution alone. This is an operational definition for design, not a claim that the literature has settled on a single meaning. Integrated STEM is a related but separate literature. It covers combinations of science, technology, engineering and mathematics, whereas the present focus can also involve integration among scientific disciplines and selected contributions from geography.

A disciplinary-necessity test is proposed: remove one contribution and ask whether the explanation, uncertainty judgement or defensible decision changes. If removal merely shortens a presentation, the contribution may be thematic enrichment rather than substantive integration. This does not make enrichment educationally worthless. It makes the claim about integration more precise. The test is stricter than counting subjects, but it does not require equal lesson time for every discipline or the inclusion of every STEM component.

A second, complementary test concerns evidence dependency. A later claim should depend on interpreting evidence generated or examined earlier, and that dependency should cross a disciplinary boundary at an identified point. For example, a low optical-scattering signal cannot support a broad water-safety claim without chemistry and biology specifying what that signal fails to measure. The disciplines are connected through a change in the warrant for a conclusion; appearing on the same worksheet is not enough. Both tests are original design heuristics requiring validation, not established psychometric scales.

Interdisciplinary assessment should retain disciplinary grounding while examining the quality of integration and reflection on its limits. Boix Mansilla and Duraising (2007) offer a relevant assessment perspective, although their higher-education setting does not validate its direct use with younger pupils. The framework developed here adapts the general assessment problem to junior secondary tasks. Students need enough disciplinary understanding to make a connection scientifically defensible. They also need enough awareness of evidence limits to avoid treating every connection as equally useful.

2.2 Modelling and shared artefacts

Modelling provides a potential meeting point because different disciplines can contribute constraints on a common representation. Hallström, Norström, and Schönborn (2023) investigate authentic STEM education through modelling using an international Delphi study. The evidential status is expert consensus, not a trial of student learning. In the present design, modelling is used to expose relationships among the source of water, possible substances or organisms, measurement procedures, treatment functions and the claims that can be supported.

The common artefact is an evidence-based decision dossier rather than a display assembled from unrelated subject sections. Its core pages are a causal map, an evidence ledger, a comparison of alternatives and a revision record. Akkerman and Bakker's (2011) account of boundary crossing and boundary objects informs this coordination role. A

shared document does not become productive merely because several teachers can access it. It must allow different disciplinary contributions to be interpreted together while keeping their distinct meanings and limitations visible.

For example, “clear” describes an appearance under a viewing or measurement condition, and “low dissolved-ion content” concerns a different property. “Microbiologically acceptable for a specified use” requires evidence that appearance alone cannot supply. Students are asked to retain these distinctions within one argument. The dossier thus supports a translation problem: how does a finding produced under one set of methods constrain a claim evaluated under another?

2.3 Teacher coordination and the visibility of learning

The feasibility of integration depends on more than curriculum wording. Margot and Kettler (2019) review teachers’ perceptions of STEM integration and identify the importance of attending to instructional and organisational conditions. Wu, Yang, Zhou, Xia, and Liao (2024) examine interdisciplinary teaching abilities in a meta-analysis. Their findings do not warrant a universal causal claim that a particular workshop will improve pupils’ learning. They do support treating teacher capability and professional support as design conditions rather than assuming that a new timetable is sufficient. A systematic review of primary and secondary teachers’ interdisciplinary competence points in the same direction, describing teacher education programmes and collaboration models as the main routes to building that competence (Wang & Sang, 2024). In a Chinese junior high school, Wei (2020) documented how six science teachers from different subject backgrounds developed professionally while implementing an integrated science curriculum. Tan and Wei (2025) show that science teachers interpret STEM education through differing curriculum ideologies, which shapes what “integration” comes to mean in their classrooms.

Instructional approach also matters. Halawa, Lin, and Hsu (2024) review instructional design in K–12 STEM education, documenting a range of design approaches rather than a single model applicable to every context. The present unit combines short disciplinary teaching episodes, guided investigation, comparison and revision. It is not premised on the idea that open-ended activity should replace explicit teaching. When pupils lack the knowledge needed to interpret evidence, a targeted explanation or worked example is an enabling component of integration.

The unit also separates engagement from learning. Hiwatig, Roehrig, and Rouleau (2024) investigate variation in integrated STEM implementation and attitudinal change. Their outcome focus cannot be substituted for evidence of disciplinary understanding or transfer. Accordingly, student enjoyment and perceived relevance are proposed as secondary evaluation outcomes. Individual explanatory performances remain central to judging whether pupils can use the disciplinary relationships that the project is intended to develop.

3. Conceptual Design Method

3.1 Purposive synthesis and its limits

The method is a literature-informed conceptual synthesis followed by construction of a worked instructional case. Following the methodological concern of Jaakkola (2020), the role of each theoretical contribution is made explicit rather than presenting a collection of references as a new framework. The synthesis connects four strands: definitions and quality of integration, disciplinary and interdisciplinary assessment, classroom enactment and teacher coordination, and the design of a prospective evaluation. It does not aggregate effect sizes or rank instructional approaches.

Targeted online searches during manuscript preparation used combinations including “interdisciplinary science”, “junior secondary”, “integrated STEM projects”, “mathematics cognitive demand”, “teacher collaboration” and “interdisciplinary assessment”. The final search check was 15 October 2026. A supplementary benchmark search of Elsevier, Wiley, Springer Nature, Sage and Taylor & Francis journals on 22 November 2026 added eleven publications, and all bibliographic records were re-resolved against Crossref. Publisher pages, official documents and author-affiliated records were consulted for source details and accessible claims. Foundational works were retained alongside recent studies. Because retrieval was purposive and full-text access was uneven, this is not an exhaustive or independently screened systematic review. Claims based on an abstract or publication record are confined to what that accessible material supports.

The construction process involved selecting a question with genuine disciplinary dependencies, identifying the minimum scientific knowledge required, designing a shared assessment artefact, specifying the task sequence and checking the sequence against plausible failure cases. The two design tests were then applied to each proposed contribution. A contribution was retained as central only when it changed the interpretation of evidence or the warrant for a later claim. Table 1 identifies how the literature informed these decisions without presenting the unit as already validated.

Table 1. Literature-informed design commitments and their evidential boundaries

Source strand	Design concern	Response proposed here	Boundary of the claim
English (2016); Kelley and Knowles (2016); Roehrig et al. (2021)	Subject labels do not fully specify the nature of integration.	Apply disciplinary-necessity and evidence-dependency tests to the common question.	The two tests are original heuristics, not validated measures.
Boix Mansilla and Duraising (2007); Hallström et al. (2023)	Integration should preserve disciplinary meaning and support a common account.	Use one causal model and decision dossier with distinct evidential warrants.	Higher-education assessment and Delphi consensus do not establish junior-secondary effects.
Forde, Robinson, Ellis, and Dare (2023)	Mathematics can be present without demanding substantial mathematical reasoning.	Make variation, comparability and trade-offs consequential to the decision.	The selected mathematical tasks require local curriculum checking.
Margot and Kettler (2019); McLure et al. (2022); Wu et al. (2024)	Enactment and teacher support are part of the problem, not background detail.	Specify coordination responsibilities and alternative timetable pathways.	A feasible pathway must be demonstrated locally.

3.2 Analytic checks and the status of the outputs

Three negative cases were used as analytic checks. In the thematic case, every subject produces a separate page and the final recommendation is unchanged when one page is removed. In the craft case, pupils can achieve a successful-looking product through trial and error without explaining its behaviour. In the decorative-mathematics case, calculations are performed but do not affect interpretation or choice. The proposed unit is structured to make each failure visible. These constructed cases are analytical devices, not observations of teachers or pupils.

Sandoval's (2014) distinction between design features, mediating processes and outcomes informs the prospective conjectures. The design features here include dependency prompts, a shared evidence ledger and an explicit revision requirement. Anticipated processes include detecting incompatible claims and using one discipline to qualify an inference made from another. Improved interdisciplinary reasoning is a possible outcome to test, not a result already established. In keeping with educational design research's concern for relevant and robust knowledge, feasibility and mechanism testing are specified before wider evaluation (McKenney & Reeves, 2025).

4. Design Strategies and Worked Unit

4.1 Frame a bounded question rather than an unlimited theme

The proposed unit asks: "Why can apparently clear rainwater still lack the evidence needed for a proposed use, and what can our model investigations establish?" The school setting makes the problem concrete without asking pupils to solve the school's actual water-management needs. Twelve periods of approximately 45 minutes are allocated as a planning scenario. The main scientific constructs are separation of suspended and dissolved matter, measurement as a selective representation, and the difference between absence of visible contamination and evidence about hazards. Mathematical comparison connects the constructs to a justified judgement.

The boundary of activity is explicit. Pupils do not drink, taste or use any water from the project, do not culture microorganisms, and do not handle unknown environmental water. Optional demonstrations use only teacher-approved, known materials under school procedures; supplied images and data can replace all wet activities. The synthetic filter examples evaluate limited model performance, not real-world potability or suitability for irrigation or any other use. The World Health Organization's drinking-water guidance treats safety through a broader risk-management approach; appearance alone is not an adequate basis for a safety conclusion (World Health Organization, 2026).

The unit does not require every science discipline to receive equal time. Chemistry and biological uncertainty are central to evaluating what "clean" might mean, while measurement and mathematical reasoning are necessary for comparing the supplied evidence. Geography can contribute a simple catchment pathway if it changes the account of possible inputs. Engineering contributes comparison and revision of model designs, but construction is optional. This prioritisation guards against adding activities merely to satisfy a checklist of subjects.

4.2 Map the dependencies before selecting activities

Table 2 is the unit's evidence-dependency map. Each row identifies a disciplinary contribution, the claim it can support and the later judgement that depends on it. The teacher asks pupils "What did this subject contribute?" and then "Which conclusion would become less defensible without it?" A disciplinary explanation may enter in a short teaching episode and be reused repeatedly. Time spent on a subject is not the proposed measure of its intellectual importance.

Table 2. Evidence dependencies in the school rainwater unit

Contribution	Evidence or representation	What it can establish	How it constrains the joint judgement
Physics and measurement	A fixed optical setup, reference sample and repeated readings.	A comparable change in the defined optical signal, within the limits of the protocol.	An apparent improvement cannot be accepted if lighting or geometry changed.
Chemistry	Particle account of suspended matter and dissolved ions; supplied conductivity evidence.	Why a particle-removal process need not remove dissolved substances.	Reduced cloudiness cannot be equated with removal of all chemical constituents.
Biology	Teacher-supplied account of microorganisms and missing microbial evidence.	Why organisms or their absence cannot be inferred from visual clarity.	A safety claim remains unsupported even when optical performance improves.
Mathematics	Repeated values, units, averages, ranges and comparisons against stated criteria.	How consistency, throughput and resource demands change a model-design choice.	A favourable average alone may conceal instability or an unfavourable trade-off.
Geography and design, where relevant	A simplified catchment pathway and design-function map.	Which assumptions about inputs or treatment functions require checking.	The scope of a conclusion must match the represented source and intended decision.

A common causal map records these relationships with labelled arrows rather than nouns in boxes alone. “Suspended particles affect the optical signal under this measurement setup” is more informative than an arrow simply connecting “physics” and “water”. Students mark whether an arrow represents supplied knowledge, a model assumption or evidence generated in the activity. Conflicting arrows become a focus of discussion. For example, “clear appearance implies no dissolved matter” conflicts with the particle explanation of a solution and should be revised.

The map is not expected to represent every possible contaminant or process. Its adequacy is judged against the bounded classroom question. This supports disciplinary depth while making omissions visible. The National Research Council (2012) framework’s emphasis on practices, core ideas and crosscutting concepts is compatible with this use of models and evidence. The present activity nevertheless remains a local design that must be mapped onto the actual syllabus rather than declared universally curriculum-aligned.

4.3 Sequence disciplinary preparation, integration and revision

The twelve-lesson sequence in Table 3 contains alternating moments of disciplinary preparation and joint reasoning. Integration is not postponed until a final poster. Lessons 3–4 already require pupils to reconcile visual appearance with a particle account; lessons 7–8 require them to reconsider a design comparison when variation and missing evidence are introduced. Short individual responses precede group discussion so that the eventual dossier does not conceal who has understood the relationships.

Table 3. Proposed twelve-lesson implementation sequence

Lessons	Task and disciplinary emphasis	Shared output	Individual checkpoint
1–2	Interrogate a fictional school proposal; distinguish observations, assumptions and questions.	Initial causal map and evidence-needed list.	Explain why “clear” and “safe” are different claims.
3–4	Study suspended versus dissolved matter and what filtration is intended to do; use approved demonstrations or supplied representations.	Revised particle explanation linked to the map.	Predict what one separation process can and cannot establish.
5–6	Evaluate an optical measurement protocol and biological information; identify uncontrolled comparisons and unmeasured properties.	Evidence ledger with method and limitation columns.	Identify a confound and an unsupported microbial inference.
7–8	Analyse the synthetic comparison dataset; compare averages, ranges, throughput and resource credits.	Alternative-design comparison with conditional preferences.	Justify why a mean or a single best reading is insufficient.
9–10	Prepare a bounded recommendation; exchange a critique that challenges one cross-disciplinary dependency.	Decision dossier and evidence-linked revision record.	Explain which disciplinary contribution changed a judgement.
11–12	Defend the dossier, complete a new-context transfer task and reflect on the model boundary.	Final dossier distinguishing supported claims from missing evidence.	Independent explanation of an unfamiliar clear-liquid case.

Teaching episodes are deliberately short and consequential. An explanation of dissolved ions immediately supports a critique of the claim that all contaminants are visible. A discussion of measurement control is followed by a comparison in

which the illumination changed. A lesson on microorganisms is not followed by culturing; it supplies the reason that an optical result cannot establish their absence. The return to the joint question is essential. Without it, even accurate disciplinary teaching would remain a parallel collection of lessons.

Mathematical demand is protected by asking pupils to choose and justify a comparison instead of calculating every available statistic. Forde et al. (2023) found that mathematics in the integrated STEM lessons they examined was often of low cognitive demand. The present design responds by making variation and choice of criteria affect the recommendation. Calculators may support computation, but the assessed reasoning concerns what a comparison means and whether the available evidence warrants it, not speed of arithmetic.

4.4 Worked analytical example using synthetic teaching data

Table 4 is a constructed dataset for instruction. The numbers were authored to create contrasts and are not measurements from a school, filter experiment or published study. In the fictional setup, each design receives the same prepared model input at the same temperature. The input optical index is defined as 100 and conductivity as 600 microsiemens per centimetre. Lower optical-index values mean a lower signal associated with cloudiness under the stipulated fixed setup; the index is not a calibrated turbidity unit or a health indicator. Three fictional repeat runs are shown for each design.

Table 4. Synthetic teaching-only comparison; not empirical water-quality results

Model design	Output optical index: three runs	Output conductivity (µS/cm): three runs	Flow (mL/min): three runs	Resource cost (fictional credits)
A	24, 27, 25	598, 604, 601	40, 42, 38	2
B	14, 43, 20	605, 600, 598	70, 68, 72	3
C	18, 17, 19	601, 602, 599	15, 14, 16	5

Note. All values are invented for this educational vignette. The fictional conductivity instrument has a stated uncertainty of ±10 µS/cm. No microbial or other chemical safety results are supplied. The data cannot certify any real water or treatment device.

The mean optical indices for A, B and C are approximately 25.3, 25.7 and 18.0, respectively, while the observed ranges are 3, 29 and 2 index units. Design B therefore illustrates why a particularly favourable single reading does not establish consistently better performance. Design C has the lowest mean optical index in this constructed comparison, but it also has the lowest flow and highest fictional resource cost. Design A and B have similar means but markedly different observed variation. Three runs are insufficient for broad reliability claims; the exercise concerns interpretation of the supplied pattern, not statistical proof of design superiority.

The conductivity readings remain close to the fictional input value relative to the stated instrument uncertainty. The appropriate classroom conclusion is that these readings do not demonstrate a reduction in conductivity. This is not proof that every dissolved species is unchanged, and conductivity alone does not identify chemical composition or safety. The chemistry explanation is therefore needed to prevent an optical improvement from being interpreted as comprehensive purification. Biological reasoning adds a separate limitation: no microbial evidence is supplied, so absence of microorganisms has not been established.

Pupils may conditionally prefer C when optical consistency is the sole model-performance aim, or A when the stated classroom criteria place more importance on resource use and throughput. Such preferences require explicit criteria and acknowledgement of the tiny constructed sample. There is no warranted overall winner for real water safety. An indiscriminate weighted score would be particularly misleading if unknown safety properties were assigned a convenient neutral value. The design therefore introduces a safety-evidence gate: missing evidence cannot be compensated for by low cost, high flow or a clear appearance.

This gate is a rule for the classroom argument, not an operational water-treatment standard. Any actual reuse decision would require appropriate authorities, applicable requirements and evidence outside this unit. The relevant educational point is logical: evidence about one property cannot simply substitute for evidence about another. The principle is consistent with the risk-based approach of World Health Organization (2026) guidance, while the detailed classroom comparison and numerical values remain original illustrative materials.

4.5 The decision dossier as an integration requirement

The final dossier contains four connected components. The causal map states the disciplinary relationships. The evidence ledger records where each item originated, what procedure or assumption supports it, and what claim it cannot justify. The comparison explains how criteria and uncertainty affect a model-design preference. The revision record identifies a claim changed after critique and the evidence responsible. These components may be completed on paper; a digital platform is not a prerequisite for integration.

A dossier is not considered substantively integrated merely because it contains all four headings. Pupils must explain at least one instance in which a disciplinary contribution qualified a claim from another source. For example: “We originally chose B because one optical reading was lowest; the repeated values changed our view of consistency, and the

chemistry and biology evidence prevented us from calling any output safe." This is an authored example of the intended form of reasoning, not a quotation from a participant.

Peer critique is similarly focused. Instead of commenting generally on presentation, each reviewing group challenges one dependency. It asks whether measurements are comparable, whether a process addresses the property claimed, or whether an unmeasured property has been assumed. The producing group must respond through a revision or a justified defence. A requirement to revise does not mean that every criticism is correct; evaluating an unsuitable criticism is itself an opportunity for disciplinary reasoning.

5. Implementation Pathways and Governance

5.1 Three organisational pathways

Schools need not begin with an extended project week or simultaneous co-teaching. The core requirement is coordination of evidence and assessment at a few identified points. Table 5 offers three implementation pathways. They differ in staffing and timetable demands, not in a presumed hierarchy of educational quality. A modest science-led unit may be more coherent than a large cross-school event if its disciplinary dependencies are actually taught and assessed.

Table 5. Alternative organisational pathways for the same core design

Pathway	Organisation and minimum coordination	Core safeguard	Principal risk to monitor
Science-led unit	One science teacher leads; a mathematics colleague reviews the comparison task and rubric before teaching.	Retain explicit mathematical reasoning, biological uncertainty and individual assessment.	Other disciplines become borrowed labels without adequate content checking.
Co-planned subject sequence	Science and mathematics lessons share the question, evidence ledger and two agreed integration checkpoints.	Use compatible units, terminology, task versions and assessment expectations.	Parallel lessons drift apart or pupils fail to carry evidence across them.
Extended project period	A team coordinates longer sessions, optional model construction and a common final review.	Preserve disciplinary teaching episodes and an independent transfer task.	Logistics, display quality or group construction displace scientific explanation.

Teacher-designed units show why such agreements matter. In a three-year study of teacher teams designing interdisciplinary units for education for sustainable development, Scheie, Arntzen and Haug (2025) analysed how subject-specific, interdisciplinary and emotional aspects were balanced in practice. A minimum coordination agreement specifies the common question, the non-negotiable content distinctions, which teacher checks each evidence source, when pupils revisit the shared map and who assesses the individual task. This agreement is more actionable than a general instruction to "collaborate". Teachers also need one shared vocabulary sheet that flags rather than erases differences in terms such as model, evidence, uncertainty and safe. Where colleagues use different conventions, pupils should be helped to understand the difference instead of encountering unacknowledged inconsistency.

Time and material assumptions must be explicit. The proposed twelve periods include comparison, critique and individual transfer; they are not additional activities added to a full topic sequence without displacement. Printed datasets, prepared images and paper dossiers support a low-resource pathway. Optional sensor use may enrich discussion of measurement but should not change who can complete the central reasoning task. Resource savings are not asserted empirically here; actual preparation and lesson time must be recorded during a pilot.

5.2 Readiness, safe enactment and staged expansion

Before a pilot, the responsible teachers should complete the tasks, review scientific accuracy, check local curriculum fit and verify the institutional procedures for any practical materials. School rainwater remains a context represented through supplied information, not a substance pupils must collect or use. The safety boundary and the distinction between synthetic and empirical data must be present on student-facing materials as well as in the teacher guide. External experts can advise on the context, but their involvement does not automatically validate the learning design.

Expansion should be conditional on evidence of readiness. The first stage asks whether pupils understand the prompts and teachers can coordinate the evidence checkpoints. The next examines whether revisions actually use disciplinary relationships and whether individual assessments are interpretable. Only after these conditions are reasonably supported should additional classes or contexts be introduced. These are proposed decision gates, not numerical benchmarks proven to predict success. A school may need to simplify the question, strengthen prerequisites or revise coordination rather than scale the same design unchanged.

A concise implementation record should document scheduled and actual time, resources, task adaptations, missed integration checkpoints and reasons for those changes. Teacher collaboration can be studied through task versions and planning decisions rather than only self-reported enthusiasm. The record makes it possible to distinguish an unworkable

design from an otherwise usable design disrupted by scheduling or access constraints. It also helps prevent post hoc descriptions of an intervention that bear little relation to what pupils experienced.

5.3 Participation and inclusion

Interdisciplinary group work should not make technical ownership or linguistic confidence a prerequisite for demonstrating understanding. The proposed arrangements include individual prediction before discussion, rotating responsibilities, readable graphs, labelled diagrams and oral explanation options where appropriate. Pupils should not be asked to disclose household water practices or personal health experiences. The fictional school scenario allows discussion of a public issue without treating pupils' homes as research sites.

These arrangements are design intentions rather than evidence of equitable outcomes. A future evaluation should examine access to the reasoning work itself: who interprets data, who proposes explanations, whose criticism changes the dossier and who can explain the final judgement independently. Group participation counts and attractive displays are insufficient proxies. Where reading or language demands obstruct the intended science, supports should be adjusted and documented without quietly changing the assessed construct.

6. Assessment and Prospective Evaluation

6.1 Assess disciplinary adequacy and integration separately

Valid assessment of interdisciplinary learning remains a known weakness of the field. A review of 49 empirical studies found that assessments more often targeted monodisciplinary knowledge than interdisciplinary understanding, although one instrument did address students' interdisciplinary understanding of drinking-water safety (Gao, Li, Shen, & Sun, 2020). Concept mapping offers one complementary route; Kellinghusen et al. (2026) used semi-open concept maps to examine how 44 secondary students connected terms from different teaching environments in a climate-change unit. Assessment here addresses four dimensions: disciplinary adequacy, cross-disciplinary use of evidence, interpretation of uncertainty and reasoned revision. Table 6 provides provisional anchors for scores 1–3; score 0 indicates no relevant evidence or an account incompatible with the central task. The rubric separates a scientifically incorrect connection from a scientifically accurate but disconnected collection of facts. It also distinguishes acknowledging missing evidence from merely adding a generic phrase such as "more research is needed".

Table 6. Provisional rubric for individual interdisciplinary reasoning

Dimension	1: Fragmentary	2: Connected	3: Integrated and qualified
Disciplinary adequacy	Uses relevant terms with important conceptual errors.	Provides accurate accounts of the key disciplinary contributions.	Uses the accounts accurately while explaining their scope and limitations.
Cross-disciplinary use	Lists subject contributions without changing the central claim.	Uses evidence from more than one discipline in the same argument.	Shows how one contribution qualifies or changes another inference.
Uncertainty and comparison	Relies on one value or treats unmeasured properties as known.	Considers variation, units and at least one missing item of evidence.	Separates kinds of uncertainty and explains their consequences for the decision.
Reasoned revision	Changes wording or appearance without an evidential reason.	Revises a claim in response to a relevant criticism.	Traces the revision to evidence and explains what remains unresolved.

Note. Score 0: no interpretable evidence or a fundamentally incompatible account. The rubric is authored for this proposed unit and has not been validated.

The group dossier supports formative feedback, but the primary learning evidence is an independent response to a related new situation. One transfer item presents a clear-looking liquid after a different separation process and supplies limited measurement information. Pupils must identify what is established, what remains unknown, and which additional disciplinary evidence would change the claim. The task provides any unfamiliar context knowledge needed so that it does not inadvertently test specialist experience rather than interdisciplinary reasoning.

Assessment evidence should change instruction during the unit. A pupil who accurately describes filtration but infers microbial safety needs a task addressing the missing warrant; another separation exercise will not help. A pupil who identifies uncertainty but cannot interpret repeated values needs support with comparison. This use of evidence is consistent with the formative assessment orientation of Black and Wiliam (2009). The criterion is whether the teacher's or learner's next action changes, not whether an online quiz or rubric has been used.

Before research deployment, the rubric requires content review, cognitive interviews with age-appropriate learners, trial coding and evidence about scoring consistency. The intended interpretation should be stated in advance, following

the validity orientation of the testing standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). A common total score is not assumed to be meaningful. Reporting dimension profiles may reveal a pupil who is strong in disciplinary knowledge but weak in integration, a distinction that an overall project grade could conceal.

6.2 Conjectures, comparison and process evidence

Three design conjectures are proposed. First, a shared evidence-dependency map should increase the visibility of unsupported cross-disciplinary inferences. Second, a dossier requiring evidence-linked revision should support more substantive integration than a collection of independent subject products. Third, explicit comparison of variation and missing evidence should reduce unjustified overall rankings. Each conjecture concerns an observable process and a corresponding individual performance. Mixed results should be expected. After a co-designed integrated STEM unit built around socioscientific issues, middle school students improved in reasoning about complexity, perspective taking and the affordances of science, but not significantly in inquiry or scepticism (Ozturk & Roehrig, 2025). Improved interest, efficient group production or completion of every activity would not by themselves confirm these conjectures.

An initial feasibility study should examine whether pupils understand the fictional setting and the status of the synthetic data, whether teacher coordination is workable, and whether the intended dependencies appear in discussion and written reasoning. Only then should a comparative evaluation be designed. A useful comparator would use the same water theme, disciplinary content and approximate lesson time but organise the products in parallel without the dependency and revision requirements. This would test the proposed added value of integration more directly than comparing the unit with unrelated content or a substantially shorter lesson sequence.

Where feasible, classes or schools could be allocated prospectively to conditions. Where assignment is non-random, baseline knowledge, teacher differences and selection processes must be measured and reported, and causal interpretation restricted. A primary outcome and analysis plan should be specified before outcome inspection. Sample size should be justified through power or precision analysis that incorporates clustering; no sample size, effect estimate or recruitment event is asserted here. Immediate and delayed independent tasks would examine whether any observed performance extends beyond the shared dossier.

Qualitative evidence should trace claims across successive dossier versions and sampled classroom episodes. The unit of analysis is an evidential move, such as revising a safety inference because optical evidence is selective, rather than the mere co-occurrence of discipline names. Contradictory and unsuccessful cases should be retained. Interviews may clarify what pupils intended, but should not replace the actual task evidence. Researchers should distinguish their interpretation from the participants' explanations and examine whether coding rules can be applied consistently.

Implementation and learning evidence must be interpreted together. A weak outcome could reflect inaccessible prompts, inadequate prerequisites, missing coordination or an incorrect design conjecture. A favourable outcome could reflect additional discussion time or unusually intensive teacher preparation. Later component studies could hold those features more closely constant. Any school-based research would require the relevant institutional approvals, appropriate consent and assent, secure data handling and a clear separation between access to teaching and permission to participate in research.

7. Discussion and Conclusion

The framework offers a specific answer to a persistent design ambiguity: integration is substantive when disciplinary contributions alter the warrant for a shared explanation or decision. The disciplinary-necessity test identifies contributions that do intellectual work, while the evidence-dependency test makes the sequence of that work visible. The rainwater example illustrates the distinction analytically. Optical improvement, dissolved-substance reasoning, repeated-value comparison and biological uncertainty do more than add four subject perspectives. Together they change which conclusions can responsibly be drawn.

This account differs from a demand that all projects include all subjects, all teachers or extensive construction. Its central organisational unit is a shared evidential decision. The three implementation pathways preserve that unit while allowing different staffing arrangements. The approach also resists treating mathematics as a supporting calculator and safety as a final reminder. Mathematical variation can change the model comparison, and missing safety evidence can invalidate an otherwise attractive recommendation. These are proposed conditions for rigorous reasoning, not claims that the resulting unit is effective in every classroom.

Important limitations remain. The literature synthesis is selective, and integrated STEM evidence is not interchangeable with evidence for every form of interdisciplinary science. Some conceptual foundations derive from higher education or expert consensus rather than junior secondary interventions. The numerical vignette is synthetic and intentionally simplified; it cannot represent the variability of actual water systems. The proposed rubric and design tests are not validated, and the twelve-period schedule is a planning allocation. Teacher expertise, curriculum sequencing and local resources may substantially affect feasibility.

The framework could also become a checklist that pupils learn to imitate. Naming two disciplines or adding a sentence about uncertainty is not the intended endpoint. Independent transfer, critique of an apparently plausible claim

and examination of actual evidence-linked revisions are therefore central to evaluation. Future studies should also consider whether the design privileges verbally confident pupils, whether subject knowledge remains secure and whether coordination demands are sustainable after an initial research partnership ends.

Interdisciplinary science education in junior secondary schools should be designed around necessary relationships rather than assembled around a theme alone. This article has operationalised that principle through two design tests, a bounded worked unit, a shared dossier, alternative implementation pathways and a prospective assessment and research plan. Its contribution is a specification that teachers and researchers can scrutinise and test. The next empirical task is to determine when those design features support genuine disciplinary connection, for which learners, and under what classroom conditions.

Acknowledgement

[Authors to confirm any acknowledgements and funding, or state that the work received no specific funding.]

Ethics and Data Availability

No human participants were recruited and no classroom or environmental measurements were collected for this conceptual article. Table 4 is an original synthetic teaching dataset, not empirical evidence. All proposed school-based research is contingent on appropriate permissions and ethics procedures. The worked tasks, dataset and provisional scoring descriptions are included in the article and appendix.

AI-Assistance Disclosure

Generative artificial intelligence was used to assist literature discovery, conceptual drafting, language editing and preparation of this manuscript. The named authors must verify all sources, approve all claims and take responsibility for the final submission; the disclosure must be adapted to the target journal's policy and the actual author workflow.

Conflict of Interest

[Authors to confirm: "The authors declare no conflicts of interest."]

Appendix A. Student-Facing Task and Teacher Review Guide

Task A1: A fictional school committee claims that a model filter's output is safe because it looks clear. Use the supplied particle, measurement and biological information to evaluate the claim. Explain one conclusion the evidence supports and two conclusions it does not support. Identify which disciplinary contribution changes the committee's reasoning. Do not propose tasting, culturing or using the water. The scenario is for argument analysis, not a real school-use decision.

Task A2: Using the synthetic data in Table 4, compare designs A and B. A pupil selects B because its best optical reading is lower. Evaluate that decision using an appropriate summary and the repeated values. Then compare A and C under two different stated priorities. Explain why these comparisons do not establish that any output is suitable for drinking or another real use. A correct numerical calculation without an explanation of its relevance is incomplete.

Task A3: In a new fictional investigation, the apparent optical improvement is measured with brighter illumination after treatment than before it. Explain why the comparison may be uninterpretable, how this affects the joint claim, and what controlled evidence would be needed. The intended performance connects measurement conditions with the evidential status of a chemical or biological conclusion; it is not a request to design a water-safety certification protocol.

Teacher review guide: Before use, check the local meaning of junior secondary, prerequisite knowledge, vocabulary load, accessibility, timing and practical-material procedures. Retain the distinction between an optical index and calibrated turbidity, and between conductivity and a complete chemical analysis. Do not remove the synthetic-data label or substitute real environmental samples without an entirely separate, institutionally reviewed plan. Record any change to the common question, dependency checkpoints or individual assessment.

Expected integration evidence: A strong response explains that a controlled optical comparison concerns a selected property, that a separation process has a limited function, that repeated readings affect the strength of a design comparison, and that unmeasured biological or chemical properties remain unresolved. It identifies the consequence of those limits for the claim. These are authored scoring expectations, not observed participant findings.

References

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169. doi:10.3102/0034654311404435

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. doi:10.1007/s11092-008-9068-5

Boix Mansilla, V., & Duraising, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *The Journal of Higher Education, 78*(2), 215–237. doi:10.1080/00221546.2007.11780874

Chen, G., Yin, C., & Wang, D. (2025). Research hotspots and trends on interdisciplinary teaching and learning in China: A bibliometric analysis. *Future in Educational Research, 3*(4), 640–654. doi:10.1002/fer3.70015

Cui, Y., Yang, C., Zhou, W., Dong, Z., Guo, H., & Zhang, Z. (2026). A "Chinese approach" to STEM education innovation and development. *ECNU Review of Education, 9*(1). doi:10.1177/20965311251394054

Deng, W., Shen, X., Hao, Y., & Zhang, H. (2025). Designing integrated STEM curriculum to enhance STEM literacy among middle school students: Insights from Chinese context. *International Journal of Technology and Design Education, 35*(5), 1785–1804. doi:10.1007/s10798-025-09962-5

English, L. D. (2016). STEM education K–12: Perspectives on integration. *International Journal of STEM Education, 3*, 3. doi:10.1186/s40594-016-0036-1

Forde, E. N., Robinson, L., Ellis, J. A., & Dare, E. A. (2023). Investigating the presence of mathematics and the levels of cognitively demanding mathematical tasks in integrated STEM units. *Disciplinary and Interdisciplinary Science Education Research, 5*, 3. doi:10.1186/s43031-022-00070-1

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education, 7*, 24. doi:10.1186/s40594-020-00225-4

Halawa, S., Lin, T.-C., & Hsu, Y.-S. (2024). Exploring instructional design in K–12 STEM education: A systematic literature review. *International Journal of STEM Education, 11*, 43. doi:10.1186/s40594-024-00503-5

Hallström, J., Norström, P., & Schönborn, K. J. (2023). Authentic STEM education through modelling: An international Delphi study. *International Journal of STEM Education, 10*, 62. doi:10.1186/s40594-023-00453-4

Hiwatig, B. M. R., Roehrig, G. H., & Rouleau, M. D. (2024). Unpacking the nuances: An exploratory multilevel analysis on the operationalization of integrated STEM education and student attitudinal change. *Disciplinary and Interdisciplinary Science Education Research, 6*, 18. doi:10.1186/s43031-024-00108-6

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review, 10*, 18–26. doi:10.1007/s13162-020-00161-0

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*, 11. doi:10.1186/s40594-016-0046-z

Kellinghusen, A., Sprenger, S., Schuck, P., Orschulik, A., Vorhölter, K., & Schulz, S. (2026). Assessing students' interdisciplinary understanding of socio-scientific issues in STEM education through concept mapping. *International Journal of Science Education*. Advance online publication. doi:10.1080/09500693.2026.2685785

Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM Education, 6*, 2. doi:10.1186/s40594-018-0151-2

McKenney, S., & Reeves, T. C. (2025). Educational design research for relevant & robust scholarship. *Journal of Computing in Higher Education, 37*, 614–638. doi:10.1007/s12528-025-09456-2

McLure, F. I., Tang, K.-S., & Williams, P. J. (2022). What do integrated STEM projects look like in middle school and high school classrooms? A systematic literature review of empirical studies of iSTEM projects. *International Journal of STEM Education, 9*, 73. doi:10.1186/s40594-022-00390-8

Ministry of Education of the People's Republic of China. (2022). *Compulsory education curriculum programme (2022 edition)* [Chinese-language policy document]. Beijing, China: Author. Retrieved from https://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582343217634.pdf

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. doi:10.17226/13165

Ozturk, N., & Roehrig, G. H. (2025). Effects of an integrated STEM unit designed around socioscientific issues on middle school students' socioscientific reasoning. *International Journal of Science and Mathematics Education, 23*(5), 1493–1518. doi:10.1007/s10763-024-10517-8

Roehrig, G. H., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research, 3*, 11. doi:10.1186/s43031-021-00041-y

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36. doi:10.1080/10508406.2013.778204

Scheie, E., Arntzen, M., & Haug, B. S. (2025). Teachers' implementation of self-designed interdisciplinary curriculum units in education for sustainable development. *Research in Science Education, 55*(4), 793–815. doi:10.1007/s11165-025-10279-5

Tan, L., & Wei, B. (2025). How science teachers deal with STEM education: An explorative study from the lens of curriculum ideology. *Science Education, 109*(1), 82–105. doi:10.1002/sce.21904

Tonnetti, B., & Lentillon-Kaestner, V. (2023). Teaching interdisciplinarity in secondary school: A systematic review. *Cogent Education, 10*(1), 2216038. doi:10.1080/2331186X.2023.2216038

Wang, H., & Sang, L. (2024). Interdisciplinary competence of primary and secondary school teachers: A systematic literature review. *Cogent Education, 11*(1), 2378277. doi:10.1080/2331186X.2024.2378277

Wei, B. (2020). An exploratory study of teacher development in the implementation of integrated science curriculum. *Research in Science Education, 50*(6), 2189–2206. doi:10.1007/s11165-018-9768-x

World Health Organization. (2026). *Guidelines for drinking-water quality: Fourth edition incorporating the first, second and third addenda*. Geneva, Switzerland: Author. Retrieved from https://www.who.int/publications/i/item/9789240121225

Wu, X., Yang, Y., Zhou, X., Xia, Y., & Liao, H. (2024). A meta-analysis of interdisciplinary teaching abilities among elementary and secondary school STEM teachers. *International Journal of STEM Education, 11*, 38. doi:10.1186/s40594-024-00500-8