Review

What Is the Practical For? A Critical Review of Laboratory Work in Secondary Chemistry and a Framework for Epistemic Design

Sijin Qiu[1,*], Xinyue Luo[1] and Ruiyang Gong[1]

[1] Zhejiang Normal University, Jinhua, China
* Correspondence: Sijin Qiu, Zhejiang Normal University, Jinhua, China

Abstract: Practical work is the most expensive and most widely mandated component of secondary chemistry teaching, yet four decades of research have failed to establish a stable relationship between handling substances and understanding chemical ideas. This article argues that the difficulty lies less in the quantity or authenticity of laboratory activity than in the epistemic work that tasks require students to perform. A critical narrative review of research on school laboratories, guidance studies, and chemistry-specific work on representation is used to develop a framework of five epistemic functions: observing with purpose, modelling and explaining, coordinating representations, warranting claims, and deciding under uncertainty. The framework treats a practical task as a demand profile rather than a label, so that two activities with identical apparatus can make very different intellectual demands. Evidence is mapped onto the functions, showing that observational and mechanistic aims are comparatively well supported, whereas the justification of claims and the transfer of evidence to decisions remain weakly supported and rarely assessed. The article specifies a scaffold architecture that makes the epistemic object of a lesson explicit without reducing it to a recipe, and illustrates the argument with a two-lesson investigation of the iodine–persulfate clock reaction, for which task sequences, a representational map, an illustrative data set, and an analytic rubric are provided. The educational value of practical work, it concludes, depends on whether its tasks require students to produce and defend chemical claims.

Keywords: practical work; laboratory learning; chemistry education; epistemic cognition; scaffolding; representational competence

1. Introduction

Practical work is the part of chemistry teaching that costs the most and is defended the most. A school laboratory requires serviced rooms, consumables, storage, technical support, and timetable space that could otherwise be used for teaching. In return, curricula assign laboratory activity a central role: the Next Generation Science Standards place investigation and explanation at the centre of science learning rather than at its margins [1], and the laboratory is widely assumed to motivate students and to make chemistry real [2]. The institutional commitment is strong, durable, and expensive.

The empirical warrant for it is much weaker. Reviews across four decades converge on a similar conclusion: practical work can support learning, but the conditions under which it does so are specific and rarely evaluated [2]. The most frequently cited classroom study observed twenty-five practical lessons in English secondary science and found that students were usually successful at the observable level of a task, assembling apparatus and obtaining the expected result, while being largely unable to explain what they had observed in terms of the underlying idea [3]. A national evaluation reached a compatible conclusion: practical work was associated with affective and procedural outcomes, but its contribution to conceptual understanding was difficult to demonstrate [4]. These findings

Received: 02 September 2026
Revised: 14 October 2026
Accepted: 27 October 2026
Published: 31 October 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

are not a verdict against laboratories. They are a verdict against a particular way of organising them.

Two explanations dominate the literature, and both are partly correct. The first is that practical work serves too many purposes at once. Hodson distinguished learning science, learning about science, and doing science, and argued that a single activity is commonly expected to deliver all three while being assessed on none [5]. The second concerns guidance. Research grounded in cognitive load theory argues that minimally guided discovery exceeds the working memory available to novices, and that explicit structure is necessary before independent investigation becomes productive [6]; meta-analytic evidence supports this direction while showing that the type and amount of guidance interact with the outcome measured [7]. The recommendation to "add more structure" is therefore too coarse to guide a teacher deciding what to write on a worksheet.

This article proposes a third framing that is compatible with both explanations but more prescriptive. Practical work should be evaluated by its epistemic function: the kind of claim a task requires students to construct, and the evidence they must coordinate in order to defend it. A titration in which students follow a numbered procedure and record a burette reading makes almost no epistemic demand. The same titration, in which students decide what would count as evidence that an endpoint was overshot, and explain whether the calculated concentration is consistent with a label claim, makes a substantial one. The difference is not the apparatus; it is what the practical work is for, made operational in the questions students are asked.

Chemistry gives the argument particular urgency, because the objects it is about are not available to the senses. A student sees a colour change or a precipitate; the ions, electron transfers, and molecular distributions that explain these events are inferred. Johnstone described the resulting burden on working memory when macroscopic, submicroscopic, and symbolic descriptions must be held together [8], and later analyses have refined the triplet without removing its pedagogical force [9]. A laboratory lesson therefore never simply shows students a concept. It shows them a signal, and the concept must be constructed from the signal by an interpretation that the task may either demand or suppress.

The article proceeds in six moves. Section 2 examines why practical work so often underperforms; Section 3 develops the framework of five epistemic functions; Section 4 maps the available evidence onto it; Section 5 specifies a scaffold architecture; Section 6 applies both to a worked investigation; and Section 7 draws out the consequences for assessment, curriculum, teacher education, and research. The scope is secondary chemistry, understood as the chemistry taught to students between approximately fourteen and nineteen years of age, together with the introductory university courses that inherit the same design problems.

2. Why Practical Work Underperforms

2.1. The Effectiveness Problem

The most influential classroom study in this area observed twenty-five practical lessons across four English secondary schools and analysed each in terms of its intended objectives, the task actually set, and the talk surrounding it [3]. Students were competent at the observable level of the tasks but the conceptual level was rarely reached, because neither the written instructions nor the teacher's questioning required a move from what had been observed to what it meant. The authors did not attribute this to teachers' subject knowledge. They attributed it to the dominance of a procedural genre in which the purpose of a lesson is to get through the activity successfully.

The pattern recurs. A national evaluation of practical work in England found that it was most strongly associated with affective and procedural outcomes and least strongly with conceptual understanding, and that the strength of the association depended on how the practical was framed and followed up [4]. The relevant variable is therefore not the presence of practical work but its framing: where tasks were followed by structured

discussion requiring students to connect observation to theory, the same apparatus produced different outcomes.

2.2. Aims Confusion and the Procedural Genre

Hodson's argument that practical work conflates learning science, learning about science, and doing science remains the clearest statement of the aims problem [5]. His point was not that these aims are incompatible in principle but that they generate different task designs and different criteria of success, and that no single activity can be optimised for all of them. Asked what a practical is for, teachers often answer "several things", and the resulting design defaults to the procedural: a task that can be completed, timed, and reported, and whose success is visible in the neatness of a results table.

The procedural genre has a distinctive grammar. It opens with an aim phrased as a topic rather than a question, continues with numbered instructions that specify quantities and sequences, requires a results table whose columns are given, and closes with questions asking students to state a definition, describe a trend, and suggest one improvement. Each element is defensible in isolation, and together they produce a document that satisfies an examiner without requiring any student to make a chemical claim. The closing questions are the critical element, because they are where epistemic demand could be introduced; when they ask students to "describe the trend" and "explain why", the explanation is usually available from the textbook rather than from the data, and the data become an illustration of a conclusion already known.

This is not a failure of authenticity in the sense of resembling professional science. Removing structure in the name of authenticity can increase procedural engagement while decreasing the likelihood that the intended concept is reached, because novices lack the schemas that make open inquiry productive [6]. Chinn and Malhotra's analysis is instructive: school inquiry tasks differ from authentic inquiry not in their apparatus but in their reasoning, since authentic inquiry involves generating questions and evaluating evidence against alternatives, whereas school tasks typically require the execution of a prescribed method and the confirmation of a known result [10]. School laboratories are unlike professional ones precisely in the epistemic operations that matter most.

2.3. Guidance and the Representational Demand of Chemistry

Working memory is limited, and novices have few schemas with which to chunk information, so instruction that leaves the selection and sequencing of information to the learner imposes a load that cannot be managed [6]. In a laboratory this is compounded, because the physical environment generates load of its own: apparatus must be located, quantities measured, hazards monitored, and observations recorded, all while the chemical idea is supposed to be constructed. The evidence on guidance is substantial but not monolithic. A meta-analysis of inquiry-based science teaching found positive overall effects, with teacher-led inquiry producing larger effects than student-led inquiry [11], and a meta-analysis focused specifically on guidance found that the optimal form depended on the outcome, with procedural support helping procedural learning and prompts helping conceptual outcomes [7]. The classic demonstration concerns the control of variables strategy: children who received explicit instruction plus probe questions were more likely to acquire and transfer the strategy than children expected to discover it from repeated practice [12].

The implication is not that laboratory work should become demonstration. It is that guidance must be aimed at the operation that is actually difficult. In many lessons it is aimed at the procedure, which students would have managed anyway, while the epistemic operation of deciding what the data show is left implicit. The result is a task heavily scaffolded where scaffolding is least needed and unsupported where it is most needed.

Chemistry compounds the difficulty. Chemical explanations are built from entities that cannot be observed, and the relationship between what is seen and what is inferred must be constructed rather than perceived. Taber's analysis of the triplet emphasises that the submicroscopic level is a model, not a scale of observation, and that teaching which

treats it as a smaller version of the macroscopic world creates systematic misconceptions [9]. Movement among representations is itself a skill: Kozma and Russell showed that experts treat a diagram as a set of relationships to be interrogated while novices treat it as a picture to be recognised [13]. Ainsworth's framework explains why multiple representations help only when the task requires learners to construct the connections between them [14]. A laboratory task that presents a phenomenon without requiring a model-based interpretation is therefore not neutral: it is likely to reinforce the assumption that the phenomenon and the equation are separate things to be memorised [15].

The three explanations converge on a single design question. If aims are unclear, the intended claim has not been specified. If guidance is misallocated, the difficult operation has not been identified. If representations are unconnected, no task requires the connection. In each case the missing element is a statement of what students are supposed to establish, and of what would count as establishing it. That statement is what the next section formalises as epistemic function.

3. Five Epistemic Functions of Practical Work

3.1. Definition

An epistemic function is the kind of justification a task requires. It is specified by three things: the claim students must produce, the evidence they must select or generate, and the relation between the two that they must make explicit. Two tasks share a function when completing them requires the same form of justification, whatever the apparatus. A task in which students decide whether a precipitate is complete and defend that decision against an alternative interpretation has the same function as one in which they decide whether an endpoint has been reached, because both require an observational judgement to be justified by reference to a chemical model.

The construct is narrower than the familiar lists of laboratory aims. Lists of aims describe what a lesson is intended to achieve, such as developing manipulative skills or fostering interest, and a lesson typically has several. Epistemic functions describe what a task requires a student to do with evidence, and a lesson can be profiled across all five at once, with a distinctive pattern. Aims are recorded in planning documents; functions are visible in the questions students are asked, which makes them the more useful unit for design and research.

3.2. The Five Functions

F1: Observing with purpose. The first function is to convert a phenomenon into a signal worth recording: deciding what to measure, at what precision, with what control of confounding variables, and what an anomaly implies for the next measurement. The demand lies in justifying the measurement decision rather than in measuring. A task that supplies the variables, the range, and the instrument has removed the function while retaining the activity, and research on the control of variables strategy shows that this reasoning is not acquired without instruction [12].

F2: Modelling and explaining. The second function is to propose a submicroscopic mechanism that would produce the observation, and to state the conditions under which the proposal holds. In chemistry this means invoking particles, their arrangement, motion, and the energy changes associated with their interaction. The demand is not correct vocabulary but an account that could fail: an explanation that specifies what would be observed if the mechanism were different, which is where progression in chemical thinking is most visible [16].

F3: Coordinating representations. The third function is to move among macroscopic description, particulate model, and symbolic notation in a way that preserves consistency. A student who writes a balanced ionic equation, draws a particle diagram of the same solution, and describes the visible change has produced three representations of one system, and the work lies in making them say the same thing. The chemistry-specific difficulty is that the symbolic level is highly compact, so a single formula encodes

stoichiometry, charge, and state in a form that novices can manipulate without understanding [15].

F4: Warranting claims. The fourth function is to evaluate whether the evidence supports the claim, and to consider what else could explain the same data. It includes reasoning about uncertainty, identifying systematic error, and recognising when a result is consistent with more than one model. It also includes the social dimension of justification: presenting a claim to an audience that may challenge it, and answering the challenge with evidence rather than authority. Such norms must be established deliberately, since students do not arrive expecting that claims require defence [17] [18],, and design studies show that prompts making the criteria for a good explanation explicit can be built into inquiry without turning it into a procedure [19].

F5: Deciding and acting. The fifth function is to use evidence to choose among courses of action, whether selecting a procedure, choosing between candidate materials, or judging whether a result is fit for a purpose. In school chemistry it is often reduced to "suggest an improvement", which invites a generic answer about repeating the experiment. Understood properly, it requires the student to specify a decision, the evidence that would inform it, and the cost of being wrong. Systems-thinking approaches make the function explicit by asking students to trace the consequences of a chemical choice through coupled processes [20]. It has the weakest empirical base in school settings and is the function most closely tied to the civic justification for teaching chemistry at all.

3.3. Functions as a Demand Profile

The five functions are not stages, and a well-designed practical need not exercise all of them. They are dimensions along which a task can be profiled. Consider three tasks on reaction rate. In the first, students time the dissolution of magnesium ribbon in three concentrations of acid and plot the results; the profile is dominated by F1. In the second, students receive the same data and must explain the trend through collision frequency and draw particle diagrams; the profile shifts to F2 and F3. In the third, students receive the data plus a textbook claim that rate doubles when concentration doubles, and must decide whether the data support the claim and what further measurement would discriminate; the profile is dominated by F4. Apparatus, topic, and time are nearly identical. The intellectual demand is not.

Table 1 sets out the three ideal-typical orientations that the framework replaces, and Figure 1 gives it a visual form. The distinction is not a hierarchy of quality: a verification practical can be the right choice when a concept must be made concrete before it can be modelled, and a procedural practical when a technique is the object of instruction. The orientation should be chosen deliberately, and the dominant orientation in school chemistry is currently procedural by default rather than by design.

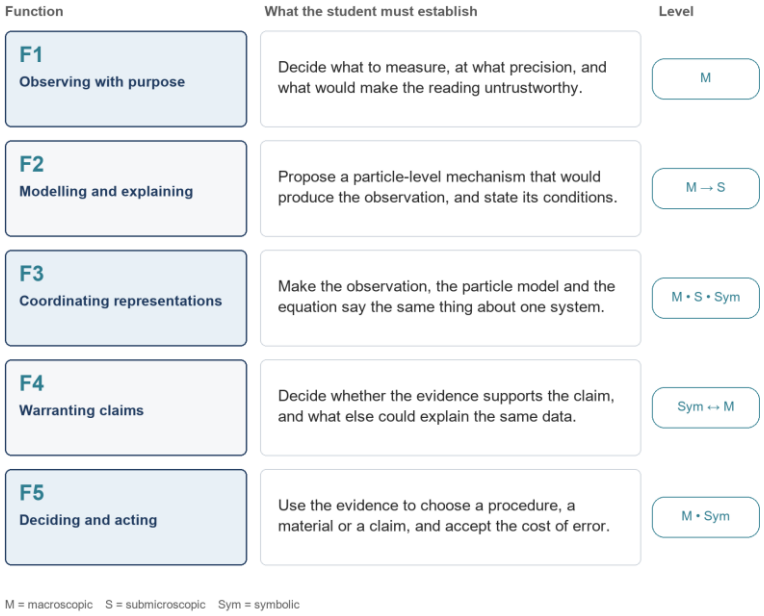


Figure 1. The Five Epistemic Functions of Practical Work. The Right-Hand Column Indicates the Description That Each Function Requires Students to Coordinate (M, Macroscopic; S, Submicroscopic; Sym, Symbolic).

Table 1. Three Ideal-Typical Orientations to a Chemistry Practical, Distinguished by the Question the Task Poses Rather Than by the Apparatus Used.

Orientation	Question the task answers	Typical instruction	Evidence students produce	Characteristic weakness
Verification	Does the concept match what we see?	"Measure the pH of these solutions and confirm that acids have pH below 7"	A table confirming an expected pattern	The conclusion is known before the data exist
Procedural	Can the operation be completed correctly?	"Titrate three samples and calculate the mean concentration"	A correct value and a neat table	Accuracy is rewarded, reasoning is not required
Epistemic	What do the data entitle us to claim?	"Decide whether the label claim is supported and state what would change your judgement"	A claim, its warrant, and an acknowledged limitation	Demands more time and more teacher subject knowledge

The figure 1 is best read as a diagnostic instrument rather than a curriculum sequence. A teacher who can locate a planned practical in two or three rows, and can say which row the assessment will actually sample, has done most of the design work that the rest of this article describes.

4. Mapping the Evidence Onto the Framework

A taxonomy earns its place by organising findings that previously appeared unrelated. Table 2 summarises the studies discussed, not to grade each function but to show where design knowledge is available and where teachers and researchers are working without it.

Table 2. Selected Evidence Mapped Onto the Five Epistemic Functions, with the Qualification That Limits Each Study's Contribution.

Study and context	Design	Function	Contribution	Qualification
Hofstein & Lunetta (2004)	Review of laboratory research	F1, F2	Establishes the laboratory's potential and the conditions for realising it	Heterogeneous outcome measures
Abrahams & Millar (2008)	Observation of 25 lessons, England	F1	Shows procedural success coexisting with conceptual failure	Few schools; science rather than chemistry alone
Abrahams & Reiss (2012)	National evaluation, England	F1, F2	Associates practical work with affective and procedural rather than conceptual gains	Survey and case design
Hofstein et al. (2005)	Inquiry-laboratory programme	F1, F4	Demonstrates gains in the quality of students' questions	Single national context
Chen & Klahr (1999)	Controlled experiments	F1	Strategy instruction with probes transfers	Not chemistry-specific
Lazonder & Harmsen (2016)	Meta-analysis of guidance	F1–F4	Guidance improves outcomes; its form matters	Heterogeneous outcome categories
Furtak et al. (2012)	Meta-analysis of inquiry teaching	F1–F4	Teacher-led inquiry yields larger effects	Broad science outcomes
de Jong et al. (2013)	Review of laboratory media	F2, F3	Identifies complementary affordances	Outcomes depend on surrounding instruction
Kozma & Russell (1997)	Expert–novice comparison	F3	Expertise is visible in the use of representations	Small samples
Kind et al. (2011)	Analysis of laboratory discourse	F4	Task features determine whether peer argumentation occurs	Limited outcome data
Sandoval & Reiser (2004)	Design study	F4	Epistemic scaffolds improve evidential quality	Students' criteria remained immature
Orgill et al. (2019)	Conceptual paper	F5	Establishes systems thinking as an organising aim	Limited school-level evidence

Functions concerning observation and mechanism are supported by decades of research. Reviews have consistently found that practical work develops the ability to handle apparatus, follow procedures, and record observations, and that these outcomes are achieved reliably even when conceptual outcomes are not [2]. Inquiry-laboratory programmes show that questioning skills can be taught deliberately, with students improving in the quality rather than only the quantity of the questions they ask [21], and controlled experiments show that the reasoning behind a measurement decision transfers only when it is explicitly taught [12]. The evidence on mechanism concerns the medium more than the task: comparisons of physical and virtual laboratories find complementary advantages, with virtual environments superior for making invisible entities visible and physical ones for equipment handling and the unexpected [22]. Where simulation studies report weak conceptual outcomes, the task profile was usually dominated by F1.

Representation is well described but weakly intervened upon. Expertise is visible in how representations are used rather than in whether they are understood [13], and multiple representations support learning only when tasks require learners to construct the mapping between them [14]. The most practical intervention evidence concerns what happens before the session: pre-laboratory activities that engage purpose, theory, and procedure reduce in-session cognitive load and free capacity for interpretation, and recent syntheses emphasise alignment across pre-laboratory, in-session, and post-laboratory activity [23] [24],.

The gap between what is known and what is done is widest for F4. Argumentation research has established that students' capacity to justify claims improves when the norms of justification are made explicit, and that such norms are not a by-product of practical activity [17] [18],. Task features, rather than the mere presence of group work, determine whether peer argumentation occurs [25], and design studies show that epistemic scaffolds improve the evidential quality of students' explanations even when their criteria remain less sophisticated than designers hope [19].

F5 has the weakest base in school chemistry, for a structural reason: decisions about what to do with evidence usually lie outside the laboratory lesson. Systems thinking has been proposed as the frame that would make such decisions central, on the argument that chemistry education should prepare students to reason about coupled technical, environmental, and social systems rather than isolated reactions [20], but classroom evidence remains limited and is mostly drawn from higher education.

The asymmetry is not accidental. F1 and F2 can be measured with written tests administered after the lesson, which suits research and examination systems, whereas F4 and F5 require the collection and analysis of students' reasoning, which is expensive. Assessment practice has followed the same logic, with the result that the functions hardest to assess are also the least likely to be taught [26].

5. A Scaffold Architecture for Epistemic Work

5.1. The Design Principle

The scaffolds described here follow one principle: support should specify the form of the reasoning required without supplying its content. A prompt asking "what would you expect to see if the mechanism were different?" specifies a form. A prompt asking students to explain that particles collide more often supplies content and removes the work. The distinction is more useful than the continuum from structured to open inquiry, because it separates the amount of support from its target. A highly structured task can still make a substantial epistemic demand if the structure is attached to the procedure rather than to the conclusion.

The principle answers a common objection. Teachers sometimes resist epistemic tasks because they appear to take time from content coverage. The evidence on guidance suggests the opposite when support is well targeted: prompts that reduce the search space free working memory for the reasoning itself [7]. What is expensive is not the scaffold but the feedback, since a teacher who asks students to justify a claim must be able to recognise a good justification and respond to a poor one.

5.2. Across the Three Phases

Before the session, scaffolds have two jobs: to make the theoretical model available, and to create a commitment that the data can later test. The first is supported by evidence that pre-laboratory activity reduces in-session load [23]. The second is less often designed for. A prediction committed to writing before the measurement creates a conflict when the data disagree, and that conflict is the occasion for the reasoning the lesson is supposed to teach. Predictions that are vague, or that students know will not be checked, produce no such occasion.

During the session, scaffolds should be attached to decision points: moments at which the student must choose between defensible alternatives, such as whether the endpoint has been reached or whether a reading is anomalous. Three prompt types are

productive: a warrant ("what in your observations supports that?"), an alternative ("what else could produce the same result?"), and a discriminator ("what would you measure to tell the two apart?"). Their virtue is that they work with any apparatus. The risk is that they become a ritual, answered with a formulaic phrase; this is reduced if the prompts are answered in writing at the moment of the decision, and if at least one prompt per lesson is followed by a whole-class comparison of two different answers, so that the criterion becomes the quality of the warrant rather than the status of the speaker [17].

After the session, a more productive structure than recall questions asks for a short written argument in four parts: the claim, the evidence, the reasoning that connects them, and a limitation. The limitation is the part most often omitted and the part with the most diagnostic value, because a student who can name a limitation has understood what the evidence can and cannot support. A comparison task follows naturally: given two conclusions drawn from the same data, one of which overclaims, students identify which is better supported and why.

5.3. Fading and Progression

Scaffolds are useful only if they are removed. The classic account treats fading as constitutive rather than optional [27], and the same applies here: a prompt that is permanent becomes a template filled in without thought. A progression might begin with a fully specified argument structure, move to a list of the four components without sentence stems, then to a single instruction to "justify your conclusion", and finally to a task in which the requirement is implied by the genre of the report. Because a single lesson does not provide enough occasions for withdrawal, fading should be planned across a sequence of practicals, so that each function is introduced with full support, practised with reduced support, and finally required without support in an unfamiliar context. Figure 2 and Table 3 set out the resulting architecture.

Table 3. Scaffold Types by Phase, with the Epistemic Function Targeted, an Example Prompt, the Artefact That Makes the Reasoning Visible, and a Route for Fading.

Scaffold	Phase	Function	Example prompt	Artefact	Fading route
Prediction with a contradiction condition	Pre-lab	F2, F4	"State what you expect and one observation that would prove you wrong"	Written prediction	From three sentences to a single claim
Representation priming	Pre-lab	F3	"Sketch the particles in 0.1 M and 1.0 M acid before you begin"	Annotated sketch	From sketch to selecting a given representation
Decision-point prompt	Enactment	F1, F4	"What in your observations tells you the endpoint is genuine?"	Justified decision	From explicit prompt to unprompted annotation
Alternative-explanation prompt	Enactment	F4	"What else could make the solution turn blue at this point?"	Short written alternative	From two alternatives to independent critique
Discriminating test	Enactment	F1, F5	"What would you measure to distinguish the two explanations?"	Proposed measurement	From proposal to executed check
Argument paragraph	Post-lab	F2, F4	"Write the claim, the evidence, and	Structured paragraph	From template to continuous prose

			the reasoning that links them"		
Limitation statement	Post-lab	F4, F5	"State one thing your data cannot show"	Written limitation	From one limitation to a ranked set
Comparing two conclusions	Post-lab	F4	"Which of these two conclusions is better supported, and why?"	Comparative judgement	From guided comparison to peer review

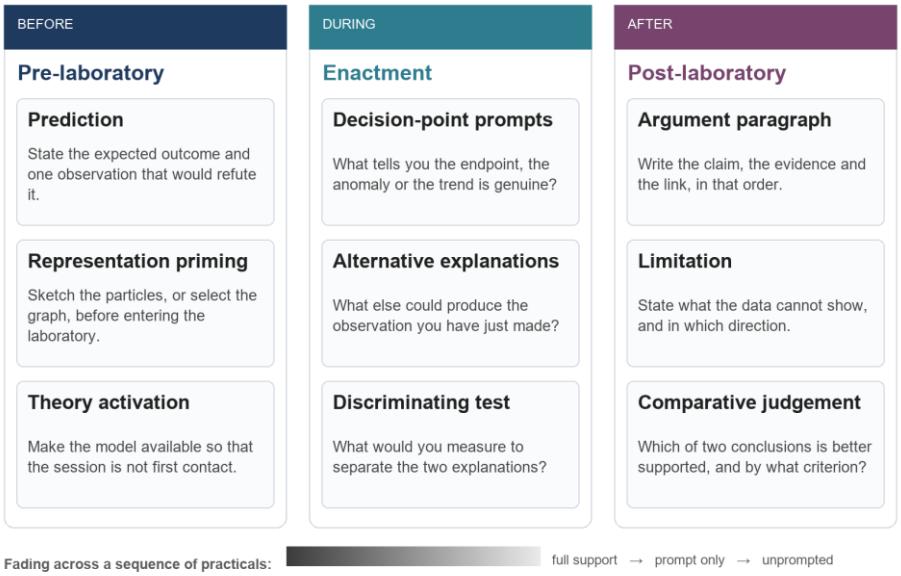


Figure 2. Scaffold Architecture Across the Three Phases of a Practical Lesson, with the Fading Intended Across a Sequence of Practicals Rather Than within One Lesson.

6. Worked Example: The Iodine–Persulfate Clock Reaction

6.1. Why This Reaction

The reaction between peroxodisulfate and iodide suits epistemic design because it has three properties that most school practicals lack. Its signal, the sudden appearance of a blue colour, is temporally sharp and easy to time. The chemistry behind the signal involves a fast secondary reaction whose role is not obvious from the equation, so the link between what is seen and what is happening requires an argument. And the temperature dependence of the time can be converted into an activation energy, which makes a model testable against data rather than merely illustrated by them. The investigation suits students aged sixteen to eighteen and occupies two lessons.

6.2. The Chemistry Students Must Reconstruct

The oxidation of iodide by peroxodisulfate is slow at room temperature and follows second-order kinetics overall, first order in each reactant:

$S_2O_8^{2-}(aq) + 2\ I^-(aq) \rightarrow 2\ SO_4^{2-}(aq) + I_2(aq)$

The iodine produced would give a gradual colour change with starch that is difficult to time. The clock is created by adding a small, known amount of thiosulfate, which reduces iodine back to iodide rapidly:

$I_2(aq) + 2\ S_2O_3^{2-}(aq) \rightarrow 2\ I^-(aq) + S_4O_6^{2-}(aq)$

Because the second reaction is much faster than the first, iodine does not accumulate while thiosulfate remains. The blue starch–iodine colour appears only when the thiosulfate has been consumed, at which point the amount of iodine produced is fixed by

the initial amount of thiosulfate, independently of temperature. The measured time therefore corresponds to a constant extent of reaction, and its reciprocal is proportional to the initial rate. This is the crucial inference, and it is exactly the kind that a procedural worksheet states rather than requires.

Three consequences follow for design. The proportionality between 1/t and the initial rate holds only if thiosulfate is in excess relative to the iodine produced during the induction period, and only if temperature is constant throughout. Constancy of temperature is questionable at the start of a run, since the mixture must equilibrate with the bath, and the resulting systematic deviation has a predictable direction. Finally, the activation energy obtained from ln(1/t) against 1/T is only as good as the assumption that the rate constant, rather than the observed time, is what varies with temperature (As shown in Figure 3).

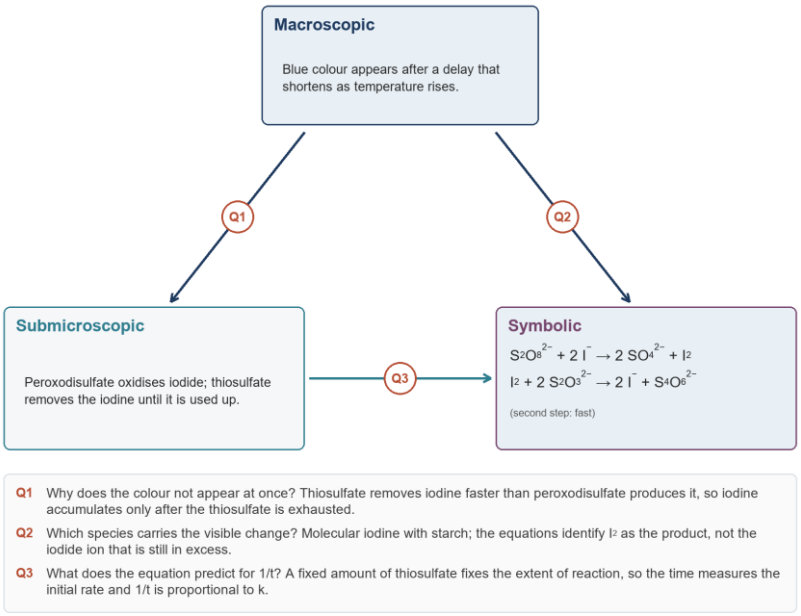


Figure 3. Representational Map for the Iodine–persulfate Clock. The Linking Questions in the Centre Column Are the Epistemic Work the Task Is Designed to Require.

6.3. Task Sequence

In the first lesson, students receive the two equations, 0.10 M potassium iodide, 0.050 M potassium peroxodisulfate, 0.010 M sodium thiosulfate, starch indicator, and a thermostatted bath. They are told that the aim is to establish how the time to the blue colour depends on temperature, and that they must decide, before beginning, which quantities to hold constant and how they will know that the bath temperature has been reached.

The pre-laboratory task asks for a prediction of the ratio between the times at 25 C and 35 C, justified from collision theory, and for one observation that would contradict it. The predictions are collected and not corrected. Students then work at five temperatures between 15 C and 35 C, and the teacher's prompts are limited to three: what tells you the mixture has reached bath temperature; what tells you the endpoint is the first persistent colour rather than the first faint tinge; and what would change if the thiosulfate concentration were halved. The third prompt exposes the logic of the clock. A student who answers that the time would double has understood that the measured time is controlled by the thiosulfate, not by the iodine; a student who answers that it would be unchanged has not, and the follow-up is a comparison of two runs rather than an explanation from the teacher.

In the second lesson, students plot ln(1/t) against 1/T, obtain a gradient, and calculate an activation energy. They compare their value with a class median and with a literature value, and account for the difference in terms of the assumptions identified above. The written product is a claim–evidence–reasoning paragraph with a limitation, followed by a comparison in which two anonymous conclusions from the same data are evaluated. The first states that the activation energy is 56 kJ mol^{-1} and that this proves collision theory; the second states that the data are consistent with about 56 kJ mol^{-1}, subject to the assumption that the mixture reached bath temperature before timing began, and that the highest-temperature point may be unreliable. Students must say which is better supported and why.

6.4. An Illustrative Data Set and What It Makes Visible

Table 4 presents data of the kind a competent pair of students might produce. Times fall from 82 s at 15.0 C to 21.0 s at 35.0 C. A regression on the first four points gives a gradient of -6.78×10^3 K, corresponding to an activation energy of 56 kJ mol^{-1}, with residuals below 0.02 in ln(1/t). Including the fifth point changes the gradient to -6.15×10^3 K and the activation energy to 51 kJ mol^{-1}, a difference of about nine per cent, and the residual at 35.0 C is −0.18, an order of magnitude larger than the others.

Table 4. Illustrative Data Set Constructed for This Article to Demonstrate the Interpretive Steps; Residuals Are Calculated from the Regression on the First Four Temperatures. The Table Reports No Empirical Findings.

Temperature / °C	Temperature / K	1/T / 10-3 K-1	Time to blue / s	ln(1/t)	Residual
15.0	288.15	3.470	82.0	−4.407	−0.015
20.0	293.15	3.411	53.0	−3.970	0.020
25.0	298.15	3.354	36.5	−3.597	0.006
30.0	303.15	3.299	25.5	−3.239	−0.011
35.0	308.15	3.245	21.0	−3.045	−0.179

That single influential point is the most valuable object in the lesson. It is consistent with a systematic error students can predict from the procedure: at the highest temperature the reaction is fast, the mixture is furthest from bath temperature when timing begins, and the early part of the run therefore proceeds more slowly than the model assumes, lengthening the measured time and lowering the apparent rate. A student who identifies this has recognised an outlier, proposed a mechanism for it, predicted its direction, and revised the estimate. A student who simply discards the point has produced a correct number without the reasoning, and the rubric in Table 5 is designed to distinguish the two (As shown in Figure 4).

Table 5. Analytic Rubric for the Two-Lesson Investigation, Aligned to the Five Epistemic Functions; Levels Describe What the Student's Text Does Rather Than How Many Correct Statements It Contains.

Criterion	Emerging	Developing	Proficient	Exemplary
Measurement decision (F1)	Records times without stating how the endpoint or bath temperature was judged	States a rule for the endpoint but not for equilibration	Justifies both decisions by reference to the procedure	Anticipates how each decision could bias the result and proposes a check
Chemical model (F2)	Describes the colour change without reference to the two reactions	Names both reactions but does not explain the delay	Explains the induction period through the relative rates	Uses the model to predict the effect of changing thiosulfate concentration

Representation (F3)	Plots the graph with errors of scale or variable	Produces a correct plot without linking it to the equations	Links the equations, the plot, and the observation consistently	Moves between the three descriptions to resolve an inconsistency
Warrant and limitation (F4/F5)	Asserts a value without evidence and states no limitation	Presents evidence but does not connect it to the claim	Constructs a claim–evidence–reasoning argument and names a limitation	Evaluates an alternative conclusion and predicts the direction of the bias

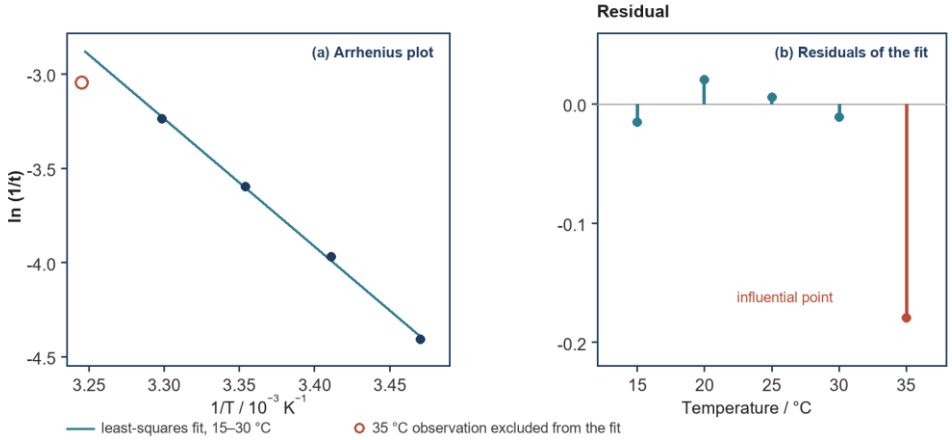


Figure 4. Illustrative Data for the Clock Reaction: (A) Ln(1/t) Against 1/T with the Regression on the First Four Temperatures; (B) Residuals, in Which the Deviation at 35 C Is an Order of Magnitude Larger Than the Others.

6.5. What the Task Makes Visible

The sequence produces evidence for each of the five functions. F1 is visible in the decision about thermal equilibration and the definition of the endpoint; F2 in the explanation of why the colour appears when it does and in the account of the influential point; F3 in the reconciliation of the two equations, the particle-level description of the induction period, and the graph; F4 in the argument paragraph and the comparative judgement; and F5, more modestly, in the choice of conditions for a reliable measurement and the recommendation about which temperatures to re-measure.

7. Assessment, Implications, and Conclusion

7.1. Assessment and Teacher Judgement

If the argument is accepted, the object of assessment is not the completion of a procedure but the quality of the reasoning a student can produce from evidence. A practical cannot be assessed epistemically by a paper that asks students to recall the expected result, because such a paper samples memory rather than reasoning; written tasks that present unfamiliar data, ask for a claim, and require a warrant can sample epistemic performance efficiently, provided the answer is not available from a textbook.

The distinction between activity and engagement is useful here. Chi and Wylie's ICAP framework distinguishes passive, active, constructive, and interactive engagement, and predicts that learning increases as students move from receiving information to constructing and co-constructing it [28]. A laboratory lesson is easily active in their sense, since students are physically doing something, while remaining passive in the sense that matters, because they follow instructions whose rationale they do not examine.

Four instruments are available, and each has a characteristic weakness. Extended written investigations sample reasoning well but are slow to mark; structured argument tasks are quicker but sample a narrow range of functions; direct performance assessment has the strongest construct validity and the weakest reliability; portfolios sample progression but are difficult to moderate. Systems relying on indirect written assessment tend to reward recall rather than the ability to investigate, whereas systems attempting direct assessment struggle with reliability [26]. Written tasks with unfamiliar data are therefore defensible for F1 and F2, whereas F4 and F5 require direct or portfolio assessment, because the quality of a warrant depends on the context in which it is produced.

Every epistemic task also generates responses that cannot be scored by a key, since a student may produce a claim that is correct for the wrong reason, or a limitation that is real but irrelevant to the data. Judging such responses requires subject-specific pedagogical knowledge. Two design features make the workload manageable: the specification of a few diagnostic responses in advance, and comparative judgement, in which the teacher ranks two or three anonymised responses against a criterion rather than scoring each against a rubric.

7.2. Implications

Practical work should be specified by its epistemic objective as well as by its content. A curriculum statement that students will "carry out a titration" specifies an activity; one that they will "justify a concentration claim in the light of the precision of their apparatus" specifies a function and makes assessment possible. The latter formulation is not more demanding of time or equipment, only of specification. Fewer practicals should also be designed more thoroughly, since the marginal practical added to a crowded schedule is unlikely to add conceptual value and may add procedural routine.

Teacher education follows the same logic. Teachers cannot prompt for reasoning they do not rehearse, and pre-service chemistry teachers are usually trained to run practicals safely rather than to anticipate the alternative explanations students will offer. A useful task is to give student teachers a protocol, ask them to profile it against the five functions, and require them to rewrite one step so that the profile shifts toward F4 without adding time or apparatus, predicting two student responses and the follow-up for each. Recognising that a student's explanation is chemically incoherent requires knowing which features carry the explanatory work, which is why generic training in questioning techniques is insufficient.

The argument applies with little modification to introductory university chemistry, where the laboratory is often the site of a first sustained encounter with uncertainty. Reid and Shah identified a tension between the skills laboratory courses claim to develop and the assessments that determine students' grades, with the result that procedural competence is rewarded and investigative reasoning is not [29].

7.3. Limitations

The framework is analytic rather than empirical: the five functions were derived from an analysis of task demands, and their independence has not been tested. Some tasks exercise several functions at once, and the boundary between F2 and F3 is porous. The evidence base is also uneven and weighted toward English-language publications and well-resourced laboratories. Finally, the worked example is a design demonstration rather than a trial: the data are illustrative, the rubric has not been used with students, and the claim that the sequence produces the intended evidence is a prediction.

8. Conclusion

Practical work in secondary chemistry is justified by an intuition that handling substances produces understanding. The research reviewed here shows that the intuition is conditional: it holds when tasks require students to move from what they observe to what they can defend, and it fails when tasks require only that a procedure be completed. The failure is not caused by insufficient apparatus, insufficient authenticity, or insufficient

enthusiasm. It is caused by a design default in which the epistemic work of the lesson is left implicit, and by an assessment system that rewards the visible completion of activity.

The framework of five epistemic functions is offered as a way of making the design choice explicit. Its practical value lies in the questions it generates. What claim must the student produce? What evidence must they select? What relation between the two must they make visible? Which function does the task exercise, and which does the assessment sample? These questions can be asked of any practical, in any topic, with any equipment, and they can be answered in a planning meeting rather than a research project.

Answering them does not require abandoning structure. It requires that structure be attached to the reasoning rather than to the procedure, and that it be withdrawn as competence develops. The worked example shows that a familiar reaction, taught with familiar equipment in two lessons, can be organised so that students must construct an argument about an influential data point and defend it against an alternative. Nothing in that example is expensive. What it requires is a teacher who knows what to listen for, and an assessment system that values what is heard. Chemistry is a discipline in which claims about invisible entities are made and defended on the basis of indirect evidence; a laboratory education that does not ask students to do the same has taught them the apparatus of the discipline without its characteristic form of thought.

Statements and Declarations: Funding: This research received no external funding. Competing interests: The author declares no competing interests. Data availability: The data set in Table 4 is illustrative and is fully reported in the table; no additional data were generated. Institutional Review Board Statement: No human participants were involved in this work. Use of generative artificial intelligence: Generative AI tools were used to assist with literature searching, reference verification, and language editing. The author reviewed and edited all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

1. National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC, USA: National Academies Press, 2012, doi: 10.17226/13165.
2. A. Hofstein and V. N. Lunetta, "The laboratory in science education: Foundations for the twenty-first century," *Science Education*, vol. 88, no. 1, pp. 28–54, 2004, doi: 10.1002/sce.10106.
3. I. Abrahams and R. Millar, "Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science," *International Journal of Science Education*, vol. 30, no. 14, pp. 1945–1969, 2008, doi: 10.1080/09500690701749305.
4. I. Abrahams and M. J. Reiss, "Practical work: Its effectiveness in primary and secondary schools in England," *Journal of Research in Science Teaching*, vol. 49, no. 8, pp. 1035–1055, 2012, doi: 10.1002/tea.21036.
5. D. Hodson, "Laboratory work as scientific method: Three decades of confusion and distortion," *Journal of Curriculum Studies*, vol. 28, no. 2, pp. 115–135, 1996, doi: 10.1080/0022027980280201.
6. P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," *Educational Psychologist*, vol. 41, no. 2, pp. 75–86, 2006, doi: 10.1207/s15326985ep4102_1.
7. A. W. Lazonder and R. Harmsen, "Meta-analysis of inquiry-based learning: Effects of guidance," *Review of Educational Research*, vol. 86, no. 3, pp. 681–718, 2016, doi: 10.3102/0034654315627366.
8. A. H. Johnstone, "Why is science difficult to learn? Things are seldom what they seem," *Journal of Computer Assisted Learning*, vol. 7, no. 2, pp. 75–83, 1991, doi: 10.1111/j.1365-2729.1991.tb00230.x.
9. K. S. Taber, "Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education," *Chemistry Education Research and Practice*, vol. 14, no. 2, pp. 156–168, 2013, doi: 10.1039/C3RP00012E.
10. C. A. Chinn and B. A. Malhotra, "Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks," *Science Education*, vol. 86, no. 2, pp. 175–218, 2002, doi: 10.1002/sce.10001.
11. E. M. Furtak, T. Seidel, H. Iverson, and D. C. Briggs, "Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis," *Review of Educational Research*, vol. 82, no. 3, pp. 300–329, 2012, doi: 10.3102/0034654312457206.
12. Z. Chen and D. Klahr, "All other things being equal: Acquisition and transfer of the control of variables strategy," *Child Development*, vol. 70, no. 5, pp. 1098–1120, 1999, doi: 10.1111/1467-8624.00081.

13. R. B. Kozma and J. Russell, "Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena," *Journal of Research in Science Teaching*, vol. 34, no. 9, pp. 949–968, 1997, doi: 10.1002/(SICI)1098-2736(199711)34:9<949::AID-TEA7>3.0.CO;2-U.
14. S. Ainsworth, "DeFT: A conceptual framework for considering learning with multiple representations," *Learning and Instruction*, vol. 16, no. 3, pp. 183–198, 2006, doi: 10.1016/j.learninstruc.2006.03.001.
15. M. M. Cooper and R. L. Stowe, "Chemistry education research—From personal empiricism to evidence, theory, and informed practice," *Chemical Reviews*, vol. 118, no. 12, pp. 6053–6087, 2018, doi: 10.1021/acs.chemrev.8b00020.
16. H. Sevian and V. Talanquer, "Rethinking chemistry: A learning progression on chemical thinking," *Chemistry Education Research and Practice*, vol. 15, no. 1, pp. 10–23, 2014, doi: 10.1039/C3RP00111C.
17. R. Driver, P. Newton, and J. Osborne, "Establishing the norms of scientific argumentation in classrooms," *Science Education*, vol. 84, no. 3, pp. 287–312, 2000, doi: 10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A.
18. J. Osborne, "Arguing to learn in science: The role of collaborative, critical discourse," *Science*, vol. 328, no. 5977, pp. 463–466, 2010, doi: 10.1126/science.1183944.
19. W. A. Sandoval and B. J. Reiser, "Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry," *Science Education*, vol. 88, no. 3, pp. 345–372, 2004, doi: 10.1002/sce.10130.
20. M. Orgill, S. York, and J. MacKellar, "Introduction to systems thinking for the chemistry education community," *Journal of Chemical Education*, vol. 96, no. 12, pp. 2720–2729, 2019, doi: 10.1021/acs.jchemed.9b00169.
21. A. Hofstein, O. Navon, M. Kipnis, and R. Mamlok-Naaman, "Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories," *Journal of Research in Science Teaching*, vol. 42, no. 7, pp. 791–806, 2005, doi: 10.1002/tea.20072.
22. T. de Jong, M. C. Linn, and Z. C. Zacharia, "Physical and virtual laboratories in science and engineering education," *Science*, vol. 340, no. 6130, pp. 305–308, 2013, doi: 10.1126/science.1230579.
23. H. Y. Agustian and M. K. Seery, "Reasserting the role of pre-laboratory activities in chemistry education: A proposed framework for their design," *Chemistry Education Research and Practice*, vol. 18, no. 4, pp. 518–532, 2017, doi: 10.1039/C7RP00140A.
24. M. K. Seery, H. Y. Agustian, F. V. Christiansen, B. Gammelgaard, and R. H. Malm, "10 guiding principles for learning in the laboratory," *Chemistry Education Research and Practice*, vol. 25, no. 2, pp. 383–402, 2024, doi: 10.1039/D3RP00245D.
25. P. M. Kind, V. Kind, A. Hofstein, and J. Wilson, "Peer argumentation in the school science laboratory—Exploring effects of task features," *International Journal of Science Education*, vol. 33, no. 18, pp. 2527–2558, 2011, doi: 10.1080/09500693.2010.550952.
26. I. Abrahams, M. J. Reiss, and R. Sharpe, "The assessment of practical work in school science," *Studies in Science Education*, vol. 49, no. 2, pp. 209–251, 2013, doi: 10.1080/03057267.2013.858496.
27. D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976, doi: 10.1111/j.1469-7610.1976.tb00381.x.
28. M. T. H. Chi and R. Wylie, "The ICAP framework: Linking cognitive engagement to active learning outcomes," *Educational Psychologist*, vol. 49, no. 4, pp. 219–243, 2014, doi: 10.1080/00461520.2014.965823.
29. N. Reid and I. Shah, "The role of laboratory work in university chemistry," *Chemistry Education Research and Practice*, vol. 8, no. 2, pp. 172–185, 2007, doi: 10.1039/B5RP90026C.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.