





Cite this: DOI: 10.1039/d5rp00422e

Diagnosing teacher candidates' reasoning on solution concentration: a theoretical analysis of definitional and procedural difficulties

Nejla Gültepe *^a and Ali Rıza Erdem ^b

This study examines teacher candidates' reasoning about solution concentration units (mass percent, molarity, and molality), with a focus on difficulties seen in definitional tasks and item-based applications. A 12-item diagnostic test informed by Cognitive Load Theory (CLT), Dual Process Theory (DPT), Conceptual Change Theory (CCT), and Representational Competence Theory (RCT) was administered to 152 teacher candidates. Additionally, semi-structured interviews were conducted with purposefully selected teacher candidates to clarify the reasoning routes underlying their response choices. Quantitative findings showed an overall accuracy of 70.29%, but performance was notably low on some items (e.g., Item 5: 32.30%), particularly when tasks required coordinating unit meaning with the correct referent and managing conversions and proportional reasoning. Error coding revealed recurring response patterns across five categories: definitional confusion (DC), unit mistake (UM), ratio-reasoning mistake (RR), superficial decision (SD), and computational mistake (CM). Interview evidence suggested that, within the formats of this instrument, some teacher candidates relied on cue-based responding and limited checking rather than explicitly grounding choices in the intended referent meaning (e.g., solution volume vs. solvent mass). Interpreted through CLT, DPT, CCT, and RCT, the findings suggest that effective instruction may require more than definitional recall and routine calculation. It may also benefit from supports that explicitly connect formula, unit, and problem context and encourage checking during problem solving. The study aims to contribute to chemistry education research by offering a theory-informed interpretation of recurring error patterns in concentration tasks based on integrated test and interview evidence.

Received 15th November 2025,
Accepted 5th April 2026

DOI: 10.1039/d5rp00422e

rsc.li/cerp

Introduction

Chemistry education is a core domain in science teaching that aims to help learners understand chemical concepts, develop problem-solving skills, and acquire scientific reasoning. Among the key topics in this field, solutions—and particularly solution concentration units (mass percent, molarity, and molality)—are central topics in both secondary and higher education curricula. Understanding concentration reasoning and calculations is important not only for laboratory practice but also for making sense of more advanced chemical phenomena such as equilibrium, reaction rates, and thermochemical processes (Johnstone, 1991; Taber, 2002).

Prior research indicates that many learners have difficulty understanding and applying concentration units (Sheppard,

2006; Naah and Sanger, 2012). Although learners may recall formal definitions, they do not always coordinate quantitative relationships involved in concentration problems. They may also fail to connect unit symbols to their contextual referents (e.g., per litre of solution vs. per kilogram of solvent) (Nakhleh, 1992; Johnstone, 1993, 2006; Pınarbaşı and Canpolat, 2003).

The science education literature suggests that learners' understanding of solution concentration is often formula-centred, while definitions, units, and referents are not always coordinated well (Sheppard, 2006; Raviolo *et al.*, 2021). Common difficulties include confusing related quantities and units, relying on surface cues or intuitive shortcuts, and struggling with proportional reasoning in dilution/mixing contexts (Staver and Jacks, 1988; Gabel, 1999; Talanquer, 2009; Raviolo *et al.*, 2021). Evidence from Turkey similarly indicates difficulties with unit conversions and operation-focused approaches in concentration tasks (Pınarbaşı and Canpolat, 2003).

Although prior research has documented that learners often struggle with concentration units, much of this work

^a Department of Mathematics and Science Education, Faculty of Education, Eskişehir Osmangazi University, Eskişehir, Türkiye. E-mail: nejlagultepe@gmail.com

^b Department of Mathematics and Science Education, Faculty of Education, Eskişehir Osmangazi University, Eskişehir, Türkiye. E-mail: aliriza.erdem@ogu.edu.tr



has primarily focused on identifying misconceptions, error types, or performance difficulties (e.g., Gabel, 1999; Sheppard, 2006; Raviolo *et al.*, 2021). While some studies in chemistry education have drawn on individual theoretical perspectives such as dual-process reasoning, representational competence, and conceptual change (e.g., Demircioğlu *et al.*, 2005; Kozma and Russell, 2005; Talanquer, 2014), integrated accounts that bring these perspectives together appear to be relatively limited.

Such a perspective is important because concentration tasks require learners not only to recall formulas or definitions, but also to coordinate unit meaning, contextual referents, and proportional reasoning across problem situations. Accordingly, the present study adopts an integrated theoretical lens to interpret recurring concentration-related errors in terms of task demands, processing tendencies, conceptual boundary instability, and symbol-referent coordination. This broader interpretive perspective may help clarify why these difficulties persist, make the diagnostic categories more conceptually specific, and provide a stronger basis for instructional design. The present study uses a diagnostic approach to examine how teacher candidates respond to concentration tasks and how recurring incorrect responses can be interpreted using reasoning patterns identified in interviews.

Theoretical framework

The present study interprets recurring response patterns and interview explanations through four complementary perspectives—Cognitive Load Theory (CLT), Dual Process Theory (DPT), Conceptual Change Theory (CCT), and Representational Competence Theory (RCT)—with a specific focus on unit-referent consistency (*i.e.*, keeping the meaning of the unit “per what?” stable across definitions, formulas, and problem contexts).

In the literature, Cognitive Load Theory (CLT) explains how performance may be affected when tasks require the simultaneous coordination of multiple interacting elements (Sweller, 1988, 2010). Dual Process Theory (DPT) distinguishes between rapid, cue-based responding and slower, more analytic checking (Evans, 2008; Kahneman, 2011). Conceptual Change Theory (CCT) concerns the stability or restructuring of key conceptual distinctions across contexts (Duit and Treagust, 2003; Posner *et al.*, 1982; Vosniadou, 2013). Representational Competence Theory (RCT) focuses on learners' ability to coordinate symbols, quantities, and contextual referents across representational forms (Gilbert and Justi, 2016; Kozma and Russell, 2005).

In this study, these perspectives are used as interpretive lenses to guide item design and to explain response patterns and interview accounts; they are not treated as variables measured. Therefore, theoretical interpretations are bounded to item demands and to the reasoning routes evidenced in the test-interview dataset.

CLT is used to characterise why some items may be more demanding than others. Concentration problems can require coordinating several elements at once (e.g., unit meaning,

referent selection, conversions, and proportional reasoning). When coordination demands increase, learners may be more likely to skip steps or lose track of the relevant referent (e.g., solution volume *vs.* solvent mass).

DPT is used to interpret differences in how decisions are made during problem solving. Some responses may reflect rapid, cue-based choices (e.g., relying on a familiar formula or a salient number), whereas other responses reflect slower checking and verification. In this study, DPT is used to describe these contrasting reasoning routes as they appear in interview explanations, without claiming to directly observe processing mode in real time.

RCT focuses on coordination between unit symbols, numerical operations, and contextual referents in text-based items. For concentration units, a key meaning question is “per what?” (e.g., mol per litre of solution in molarity; mol per kilogram of solvent in molality). RCT supports interpretation of cases where teacher candidates recall symbols (*L*, kg) but do not consistently connect them to the intended referent in context. Because the instrument does not include particulate or graphical representations, RCT interpretations are limited to symbolic-numerical-contextual coordination.

CCT is used to interpret whether key conceptual boundaries appear stable across items. Relevant boundaries include molarity *vs.* molality, solvent *vs.* solution, and amount of solute *vs.* concentration. In this study, CCT supports cautious interpretation of patterns where these distinctions shift across contexts; it is not used to claim that conceptual change occurred.

The integrative explanatory synthesis that links these lenses to the observed patterns is presented in the Discussion (see Fig. 2).

Building on the documented difficulties in the literature and the interpretive possibilities offered by the theoretical framework, the present study seeks to move beyond identifying incorrect responses and instead examine how common concentration-related errors can be explained theoretically. In this way, the research question is grounded both in the empirical problem described in the Introduction and in the conceptual lenses outlined in the theoretical framework.

Accordingly, the study addresses the following question: How can common solution concentration errors be explained theoretically?

The claims asserted in this study are limited to evidence from the diagnostic items and interview data.

Method

Research design

In this study, the diagnostic framework refers to a theory-informed structure based on CLT, DPT, CCT, and RCT, which guided both item design and interpretation of response patterns. Using this diagnostic framework, the study examined how teacher candidates reasoned about solution concentration units and interpreted recurring reasoning patterns associated with the identified error categories (Merriam, 1998; Yin, 2014).



Teacher candidates' responses to the diagnostic items and follow-up interviews were used to examine reasoning patterns and possible alternative conceptions within the scope of this instrument. The analysis went beyond correct–incorrect scoring by focusing on the reasoning reflected in response choices and clarified in interview explanations. The analysis was guided by the theoretical framework described in the previous section. This approach combined diagnostic assessment with interview evidence to provide an integrated interpretation of teacher candidates' performance. Teacher candidates were given 30 minutes to complete the 12-item test in a single session. The test was administered in regular classroom settings across four grade levels with the necessary permissions. Before the test, teacher candidates were informed about its scope, duration, and purpose. They were informed that the results would not affect their academic performance. Calculator use was not allowed. Participation was voluntary, and the study was conducted with the approval of the university's ethics committee.

Participants

The study included 152 teacher candidates enrolled in the Science Education program at the Faculty of Education, Eski-şehir Osmangazi University, Türkiye. The sample included teacher candidates from Year 1 to Year 4 of the undergraduate program (Year 1: $n = 39$, Year 2: $n = 37$, Year 3: $n = 31$, Year 4: $n = 40$); five did not report their year level. A purposive sampling strategy was used. It was confirmed that the topic of solutions and concentration units had been covered in all participating year groups. Although the study did not compare year groups, differences in coursework timing and experience may have influenced response patterns; therefore, the findings are interpreted for the sample as a whole.

Theory-based diagnostic test design

In this study, a 12-item multiple-choice diagnostic test was developed to examine teacher candidates' reasoning about solution concentration units. The test focused on definition-level

knowledge and item-based application as indicators of conceptual and procedural performance. It was designed to show how teacher candidates coordinate unit meaning, proportional relationships, and key referents such as solute, solvent, and solution while solving concentration-related tasks. The instrument addressed four cognitive dimensions across the three concentration units:

- Definition-level knowledge and symbolic labelling (Items 1, 5, 9)
- Qualitative reasoning about change (Items 2, 6, 10)
- Comparative reasoning (Items 3, 7, 11)
- Numerical operations and formula-based applications (Items 4, 8, 12)

Four items were prepared for each concentration unit: Items 1–4 focus on mass percentage, 5–8 on molarity, and 9–12 on molality. Each item was constructed based on error patterns reported in the literature (Gabel, 1999; Naah and Sanger, 2012) to identify definitional and procedural errors frequently exhibited by teacher candidates. The entire test is included in the appendices, and Table 1 presents four representative sample items. The other items (5–12) have a similar cognitive structure and are constructed in parallel patterns for molarity and molality units.

Item stems and response options were constructed from error patterns reported in prior research, with distractors designed to make both definitional-boundary confusions and process breakdowns during execution diagnostically visible within a short multiple-choice format (see Table 2 for distractor-code mappings; the full item set was provided in Appendix 1). CLT, DPT, CCT, and RCT were used as design and interpretive lenses rather than as directly measured variables. Accordingly, some items were written to increase element interactivity (*e.g.*, referent selection alongside unit conversion and proportional reasoning) so that step-skipping or context neglect under demanding conditions could be reflected in response choices (CLT). Because cognitive load was not measured directly, CLT is used here to interpret item coordination demands (not to claim participant overload or to estimate how much “load” causes

Table 1 Sample items representing the concept of mass percent concentration (% w/w)

1	Which of the following solutions has a mass percent concentration of 20%?	(A) A solution prepared by dissolving 20 grams of table salt in 100 grams of water (B) A solution containing 20 grams of dissolved salt in 100 milliliters of table salt solution (C) A solution prepared by dissolving 20 grams of table salt in 80 grams of water
2	Pure water is added to a 20% KNO_3 solution at a constant temperature. How does the mass percentage concentration of the solution change?	(A) Increases (B) Decreases (C) Remains unchanged
3	Which of the following solutions has a higher mass percentage concentration?	(A) A mixture of 20 grams of sugar and 30 grams of water (B) A mixture of 30 grams of sugar and 40 grams of water (C) A mixture of 50 grams of sugar and 80 grams of water
4	25 grams of sugar is dissolved in 100 grams of sugar solution containing 25% by mass. What is the percentage by mass of the new solution?	(A) 50% (B) 40% (C) 30%

Note. Items 1–4 represent the mass percent concentration (% w/w) dimension of the diagnostic test. The remaining items (5–8) focus on molarity, and items (9–12) on molality, each reflecting parallel cognitive structures with distinct error representations.



specific misconceptions). Some distractors were formulated to be plausible under single-cue, rapid selection (Type 1), for example by inviting reliance on a salient unit label or a visually convenient numerical relation, while interview prompts probed checking/monitoring during problem solving (Type 2) (DPT). Definition-level boundaries documented in the literature (e.g., molarity vs. molality; solvent vs. solution referents) were targeted through specific distractors (CCT). Because external visual representations were not included, RCT was operationalized as symbolic-numerical-contextual coordination within text-based items—specifically, tracking unit meanings and their referents (per liter of solution vs. per kilogram of solvent) across formulas and word-problem contexts. The absence of external representational formats (e.g., particulate diagrams/graphs) is therefore treated as a design limitation and a direction for future work, rather than as a claim that representations are unnecessary.

The number of distractors linked to a given code (e.g., CM-1) indicates diagnostic coverage across item contexts, not the prevalence or importance of that error. In this study, CM-1 (computational mistake indicating execution breakdown during problem solving) refers to a breakdown in unit-referent tracking and procedural coordination during execution steps such as conversion or substitution, rather than a simple arithmetic mistake. Because concentration tasks often involve multi-step coordination, CM-1 captures execution breakdowns that may occur during attempted analytic solving, rather than basic arithmetic or algebra errors. Core distinctions (solute–solvent–solution; solvent vs. solution referents) were probed across multiple items, and in some contexts these distinctions were reflected through related DC/UM options (e.g., Items 5 and 9). Therefore, the appearance of DC-1 (definitional confusion involving incorrect identification of solution vs. solvent) in one distractor does not mean that the solvent–solution boundary was examined only once; the same distinction was also assessed across related item contexts.

Theoretical basis of error types. The error types defined in this study were developed to capture recurring response patterns in the diagnostic test and their possible cognitive sources, as interpreted through the integrated framework. Each code reflects a recurring difficulty in maintaining conceptual boundaries, coordinating unit-referent meanings, or sustaining execution during multi-step concentration tasks.

Explanations of error types. More detailed explanations are provided for DC-1 and DC-2 because these codes are central to interpreting definition-level distinctions and referent mapping in the present study. The remaining error types are explained more briefly because they reflect similar patterns.

DC-1: definitional confusion – equating solvent and solution (mass percent context). DC-1 refers to treating the solvent amount (e.g., “100 g water”) as if it were the solution amount in a mass-percent situation (or *vice versa*), which leads to setting up the ratio with an incorrect referent/denominator. In this study, DC-1 is used as a response-pattern code linked to the DC-1 distractor design (Item 1, Option A). From a CLT perspective, mass-percent tasks can increase task demand because they

Table 2 Item–option–code matchings

Item	Option A	Option B	Option C
1	DC-1	UM-1	Correct
2	RR-1	Correct	RR-2
3	CM-1	Correct	SD-1
4	CM-1	Correct	CM-1
5	DC-2	UM-2	Correct
6	Correct	RR-1	RR-2
7	CM-1	Correct	SD-1
8	CM-1	CM-1	Correct
9	UM-2	Correct	DC-2
10	Correct	RR-1	RR-2
11	Correct	CM-1	SD-1
12	CM-1	CM-1	Correct

Note. The number of distractors mapped to each code reflects diagnostic coverage across item contexts rather than error prevalence or importance. CM-1 denotes process-level loss of definitional focus/coordination during execution (not minor arithmetic). Representational competence is operationalized as coordination among embedded symbolic (units/labels), numerical (ratios/conversions), and contextual referents within text-based items.

require tracking multiple quantities and referents simultaneously, which may make referent confusions more likely. From a CCT perspective, DC-1 reflects an unstable solvent–solution boundary, where new information may be assimilated into an incomplete schema rather than restructured (Posner *et al.*, 1982; Taber, 2002). From an RCT perspective, DC-1 involves difficulty linking symbols or quantities to their intended referents (e.g., “mass of solution” vs. “mass of solvent”) and coordinating these meanings across contexts (Kozma and Russell, 2005). From a DPT perspective, DC-1 is not defined as superficial reasoning, but it may become more visible when respondents rely on a quick choice without verifying the referent or ratio (Naah and Sanger, 2012).

DC-2: definitional confusion – solvent volume \neq solution volume. The teacher candidate evaluates only the volume or mass of the solvent as the entire solution in molarity or molality calculations. According to CLT, formula knowledge, distinguishing between units, and simultaneous execution of calculation steps overload working memory; in this case, the teacher candidate focuses only on the solvent as a mental economy. In the DPT context, the quick and intuitive inference “solvent = solution” becomes dominant, and since the analytical control process does not come into play, an incorrect decision is made. CCT attributes this error to the teacher candidate’s attempt to integrate the concept of solution with existing incomplete schemas; since conceptual transformation cannot be achieved, the distinction between solvent and solution cannot be established. RCT shows that the correct transition between volume–mass representations cannot be made and that the contextual meanings of symbolic expressions (e.g., $V_{\text{solvent}} \neq V_{\text{solute}}$) cannot be grasped. In the literature, students mixing up solvent and solution and only placing the solvent volume in formulas has been reported as a typical indicator of representational limitations (Johnstone, 1991; Naah and Sanger, 2012). Therefore, DC-2 is seen as one of the frequently reported definitional confusions related to solution concentration.



UM-1: unit error – selecting *w/v* instead of *w/w*. This error occurs when the teacher candidate uses volume (mL) instead of mass percentage (*w/w%*) in setting up the proportion. It indicates confusion about unit meaning and weak coordination between related representations. Similar mass–volume confusion has been reported in previous studies (Kozma and Russell, 2005; Naah and Sanger, 2012).

UM-2: unit error – *molarity* \neq *molality*. This error occurs when teacher candidates use the molarity formula in molality items and confuse the mass of the solvent with the volume of the solution. It often reflects rote formula selection and failure to distinguish between mass-based and volume-based units. Similar confusion between molarity and molality has been widely reported (Johnstone, 1991).

RR-1: proportion error – *failure to understand inverse proportion*. This error occurs when the teacher candidate does not recognise that concentration decreases as the amount of solvent increases and focuses only on the amount of solute. It reflects weak proportional reasoning and difficulty coordinating the relevant quantities under task demand (Sweller, 1988; Evans, 2008).

RR-2: proportion error – *assumption of constancy*. RR-2 refers to treating concentration as if it remains unchanged when one component, such as solvent amount or total volume, changes. It reflects a failure to coordinate proportional relationships (Evans, 2008).

SD-1: superficial decision – *single-cue responding*. SD-1 refers to a cue-based response in which the teacher candidate focuses on one salient feature, such as the solute amount or a numerically larger value, without coordinating key variables such as referent, ratio, and unit meaning (Evans, 2008).

CM-1: procedural/computational mistake – *execution breakdown*. CM-1 refers to errors that occur during attempted analytic execution, such as incorrect conversions, operation-order errors, or arithmetic slips. These errors may arise while applying a formula, performing unit conversions, or using molar mass where needed. From a CLT perspective, multi-step solving can increase the likelihood of such execution slips (Johnstone, 1991; Sweller, 1994).

Fig. 1 was prepared by the authors to briefly illustrate the theory-based process used in the development of the diagnostic test and the classification of error codes. The process begins with the theoretical framework, which integrates cognitive load, dual process, conceptual change, and representational competence perspectives. Based on this foundation, diagnostic items with theory-based distractors were constructed. The analysis of teacher candidates' responses and corresponding error codes (DC, UM, RR, SD, CM) provided insight into underlying reasoning patterns and areas of difficulty within the scope of this instrument. Finally, theory-driven instructional inferences were derived to guide the design of targeted remediation strategies in chemistry education.

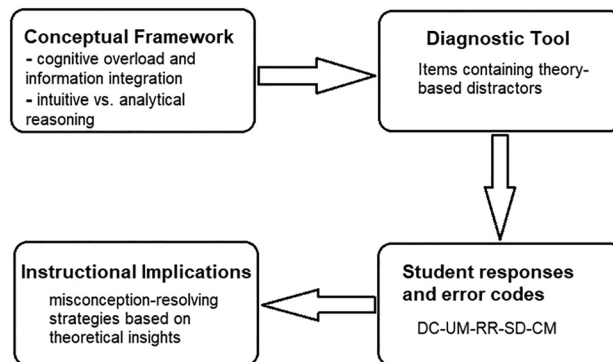


Fig. 1 Theory-based test development process and error classification model.

Test validity and expert review

The validity process was conducted in two stages. The first stage focused on linguistic and structural clarity of the item statements. A think-aloud protocol was conducted with four high school students to check readability, interpretation of wording, and clarity of instructions. High school students were used in this preliminary stage because the aim was not to validate conceptual understanding at the university level, but simply to identify ambiguous expressions or confusing words. Furthermore, the topic of solution concentration is covered in the high school curriculum at almost the same level and with similar content. This step was therefore limited to linguistic and structural refinement of the instrument. Based on this feedback, minor wording revisions were made, and linguistic consistency was reviewed with an experienced chemistry teacher.

The second stage focused on alignment between options and the intended diagnostic targets. A chemistry education expert reviewed the correspondence between distractors and the predefined error categories (DC, UM, RR, SD, CM). The expert judged most mappings as appropriate but noted that Option C in Items 3, 7, and 11 could also be interpreted as a procedural/computational error (CM-1). Following this feedback, we re-examined these options using our *a priori* code definitions and the intended reasoning pattern each distractor was designed to elicit. In our coding scheme, CM-1 is reserved for cases where an analytic procedure (*e.g.*, calculation or conversion) is attempted but an execution error occurs, whereas SD-1 captures cue-based selections made without coordinating key variables and without ratio or referent checking. Because the targeted error in these options is a failure to coordinate variables or ratio checking, the options were retained under SD-1. To minimise ambiguity, we clarified in the coding definitions that SD-1 refers to single-cue selections without ratio or referent checking, whereas CM-1 refers to execution breakdowns during an attempted calculation or conversion.

To ensure consistency with the types of errors identified by the distractors, two independent researchers conducted the coding process, during which disagreement arose particularly regarding the A options for Items 1, 3, and 5. Initially evaluated



under DC-1, these options were defined as a unique pattern where the solvent volume was confused with the solution volume, supported by the literature (Talanquer, 2009; Naah and Sanger, 2012) and interview findings, and were recoded under DC-2. This decision was theoretically and empirically grounded, and full consensus was reached among the researchers.

Using literature-based rationales alongside interview evidence helped refine the operational boundaries between codes and strengthened the content validity of distractors, consistent with diagnostic test development principles (Treagust, 1988). The test developed in this way is a theoretically and content-consistent measurement tool that reveals not only which Items teacher candidates get wrong, but also the underlying cognitive processes that give rise to these errors.

Interview process

Building upon the theoretical and empirical validation of the diagnostic test, the subsequent phase aimed to triangulate these findings through semi-structured interviews to clarify the reasoning routes underlying selected response patterns and to document how teacher candidates checked, justified, and revised their solutions when prompted. This phase increased transparency about how item selections were interpreted within this instrument (Denzin, 1978; Patton, 1999; Cohen *et al.*, 2018). Thus, quantitative and qualitative data were evaluated together, strengthening the internal validity of the study and revealing the causes of conceptual difficulties in a more comprehensive manner. The interviews were conducted one-on-one in an online environment, lasted an average of 15–20 minutes, and written consent was obtained from all teacher candidates. The protocol focused on how teacher candidates (i) interpreted definitional statements, (ii) coordinated unit meaning with the appropriate referent in contextual items,

and (iii) monitored conversions and proportional reasoning during problem solving. Sample questions included: “How did you approach solving this problem: by using a formula or by visualizing it in your mind?”, “What do molality, molarity, or mass percentage mean to you?”, “Why did you choose this option?”.

Interviewed teacher candidates were selected from the full sample of 152 using a maximum variation sampling strategy. The selected teacher candidates represented diverse error profiles, ranging from single-error to multi-error cases, and also included high-achieving candidates who nevertheless showed intuitive reasoning. Each interviewed teacher candidate was assigned a unique identifier (Tc-1, Tc-2, *etc.*), and the selection rationale is summarised in Table 3. This structure helped balance the interview sample in terms of both representativeness and diversity of error types. Interpreted together with the test data, the interview findings supported a more detailed examination of the reasoning patterns associated with each identified error type. Appendix A2 provides an interview-test map showing how selected item responses were linked to the intended code targets and how interview explanations were used to clarify the reasoning routes underlying those selections.

Data analysis

The data analysis process was conducted in two stages. In the first stage, the responses to each test item were coded as correct or incorrect, and general achievement levels were determined using descriptive statistics.

In the second stage, incorrect responses were classified according to the error categories (DC, UM, RR, SD, and CM) previously defined in detail in the section Theory-based diagnostic test design (see Table 2). At this stage, the responses given to each item were matched with the relevant error

Table 3 Interviewed teacher candidates and selection targets (primary distractor patterns)

Teacher candidate code	Dominant error type	Reason for selection
Tc-1	DC-1	19 teacher candidates made a DC-1 error; one of the 3 teacher candidates who made fewer errors was selected.
Tc-2	DC-2	124 teacher candidates made a DC-2 error; one of the 3 teacher candidates who made the same error on Items 5 and 9 and answered all other Items correctly was selected.
Tc-3	UM-1	One teacher candidate who made the UM-1 error and answered all other Items correctly was selected from the 11 teacher candidates who made the UM-1 error.
Tc-4	UM-2	Only one teacher candidate who made the UM-2 error on both Items 5 and 9 was identified and selected directly.
Tc-5	RR-1	The candidate with the fewest errors was selected from among 21 teacher candidates.
Tc-6	RR-2	One of the three teacher candidates who answered RR-2 on both Items was selected from among 35 teacher candidates.
Tc-7	SD-1	One teacher candidate was selected from among 65 teacher candidates; one of the 7 teacher candidates who answered SD-1 correctly on both Items.
Tc-8	CM-1 (w%)	One of the two teacher candidates who answered CM-1 correctly in both Items was selected from among 59 teacher candidates.
Tc-9	CM-1 (molarity and molality)	From among the teacher candidates who answered CM-1 correctly in the molar (27) and molal (62) Items, a representative was selected who chose all the CM-1 distractors in both contents.
Tc-10	Molarity (DC-2, RR-2, CM-1)	A teacher candidate who made a mistake only on Items involving molarity but answered all other Items correctly was selected.
Tc-11	Molality (UM-2, RR-1, CM-1)	The teacher candidate who made mistakes only on Items involving molality but answered all other Items correctly was selected.
Tc-12	Multiple (DC-1, DC-2, UM-2, CM-1)	A teacher candidate who made mistakes in 8 Items, exhibiting multiple error types, was selected.



category, frequency and percentage distributions were calculated (see Tables 4 and 5), and consistency patterns were identified through item-pair comparisons (definition \leftrightarrow process; qualitative \leftrightarrow quantitative) (see Table 6). Coding was performed collaboratively by the researchers, and since pre-structured error codes were used, systematic verification checks were applied instead of inter-coder reliability. Qualitative interviews were analyzed using the same code set and triangulation was achieved by comparing them with test data. This structure was used not only to identify which items teacher candidates answered incorrectly, but also the underlying cognitive mechanisms driving those errors. To increase transparency in how option selections were interpreted, interview accounts were mapped onto intended distractor-code targets to identify the reasoning routes underlying selected responses (see Appendix A2).

The analysis of interview data was conducted based on the theoretical framework that most strongly explained each error type. In this context, definitional confusion errors (DC-1 and DC-2) were evaluated at the intersection of all theories (CLT, DPT, CCT, RCT). For other error types, the analysis focused only on the theory or theories most directly explaining the relevant error to avoid repetition; in cases requiring additional explanation, reference was made to supporting theories. This approach ensured that the analyses maintained both their theoretical depth and a unique and focused interpretation structure for each error.

Findings

Response distributions and item performance

A total of 1824 item responses (152 teacher candidates \times 12 items) were analysed. Overall, 1282 responses (70.29%) were correct and 510 responses (27.96%) were incorrect. Thirty-two responses (1.75%) were missing because no option was selected. These results suggest that many teacher candidates were able to answer a substantial proportion of items correctly, indicating basic familiarity with solution concentration units, while a notable proportion of responses still reflected conceptual and procedural difficulties.

At the individual level, only one teacher candidate answered all items correctly. One additional teacher candidate achieved a near-perfect score but left one item unanswered. Fifteen teacher candidates had only one incorrect response, 27 had two incorrect responses, and 33 had three incorrect responses. In total, 80 teacher candidates achieved 75% or higher (*i.e.*, at

least 9 correct out of 12), suggesting relatively strong overall performance with clear variation across individuals.

Table 4 presents the option-level distribution for each item and helps identify recurring response patterns across the instrument. For example, the correct response rate for Item 5 was 32.3% ($f = 49$), and many teacher candidates selected the distractor that treated solvent volume as solution volume, a pattern consistent with DC-2. Similarly, Item 9 showed a relatively low success rate of 44.7% ($f = 68$), and its option distribution suggests difficulty coordinating unit meaning with the appropriate referent, which is consistent with UM-2 and, in some cases, DC-2.

The overall pattern shows relatively high success on items requiring descriptive and qualitative reasoning (*e.g.*, Items 2, 4, 6, and 8), but lower success on items requiring numerical operations and proportional reasoning (*e.g.*, Items 5, 9, and 11).

Item 2, with a correct response rate of 97.4% ($f = 148$), reflected a basic understanding of dilution and suggested that teacher candidates were successful in simple proportional reasoning. In contrast, Item 5 (32.3%) and Item 9 (44.7%) probed the molarity–molality distinction, which requires coordination of mass, volume, and symbolic representations. Responses to Item 11 similarly suggested difficulty coordinating proportional reasoning with unit conversion.

Error type-based coding and distribution

Table 5 reports the proportion of teacher candidates who exhibited at least one response consistent with each error type across the 12-item instrument ($N = 152$). Because teacher candidates could display multiple error types, percentages do not sum to 100%. Analysis of incorrect responses showed that the most frequent error types were DC-2 (solvent–solution referent confusion, 81.57%) and CM-1 (execution breakdown during attempted calculation or conversion, 70.39%). In contrast, UM-1 (unit confusion in mass-percent contexts; 7.24%) and RR-1 (inverse-proportionality breakdown; 13.82%) were less frequent.

When examining the content-based distribution, it was observed that DC-2 errors were particularly concentrated in molarity Items (63.82%), while CM-1 errors were concentrated in molality Items (40.79%) (Table 5).

Overall, the distributions suggest that some teacher candidates in this sample experienced difficulties in three areas: definitional/referent distinctions (solvent–solute, molarity–molality), ratio and proportional thinking (solute–solvent relationships), and managing the steps of a procedure (formula selection, unit conversions). In the next stage, these error

Table 4 Frequency of answers given for each option ($N = 152$)

Item	1	2	3	4	5	6	7	8	9	10	11	12
Option A	19	1	28	29	97	122	12	17	17	118	71	9
Option B	11	148	115	116	6	6	95	2	68	14	53	10
Option C	122	1	9	4	49	18	42	129	65	19	21	129
Did not answer	0	2	0	3	0	6	3	4	2	1	7	4
Total	152	152	152	152	152	152	152	152	152	152	152	152

Note. Correct responses are highlighted to indicate definitional consistency across subtopics.



Table 5 Distribution of error types by content area ($N = 152$)

Code context	Number of teacher candidates	Percentage of total teacher candidates	Overall percentage showing this error code
DC-1 (Mass%)	19	12.50	12.50
DC-2 (Molarity)	97	63.82	81.57
DC-2 (Molality)	65	42.76	
UM-1 (Mass%)	11	7.24	7.24
UM-2 (Molarity)	6	3.95	14.47
UM-2 (Molality)	17	11.18	
RR-1 (Mass%)	1	0.66	13.82
RR-1 (Molarity)	6	3.95	
RR-1 (Molality)	14	9.21	
RR-2 (Mass%)	1	0.66	23.03
RR-2 (Molarity)	18	11.84	
RR-2 (Molality)	19	12.50	
SD-1 (Mass%)	9	5.92	42.76
SD-1 (Molarity)	42	27.63	
SD-1 (Molality)	21	13.81	
CM-1 (Mass%)	59	38.82	70.39
CM-1 (Molarity)	27	17.76	
CM-1 (Molality)	62	40.79	

Note. The second percentage column shows the percentage of the full sample ($N = 152$) for each row-specific code context. The final percentage column shows the overall proportion of teacher candidates who exhibited that error code across relevant content areas. These overall percentages are not the sum of the row percentages, because the same teacher candidate could exhibit the same error code in more than one content area.

patterns were examined in depth in terms of their cognitive underpinnings, supported by findings from semi-structured interviews.

Findings from semi-structured interviews

DC-1 error-difficulty establishing the solvent-solution distinction. DC-1 was observed only in Item 1 through Option A, which was designed to capture solvent-solution referent confusion in a mass-percent context. Only 19 teacher candidates made this error. Among those who made this error, one of the three teacher candidates who made the fewest errors in other Items (Tc-1) was specifically selected for the interview based on this response pattern. During the interviews, the teacher candidate articulated the correct idea and reported marking the option quickly without double-checking (e.g., “I may have marked it directly and moved on”). This pattern is consistent with an in-the-moment monitoring lapse, possibly amplified by situational testing factors such as time pressure or anxiety, rather than a stable misconception for these cases.

All teacher candidates specifically selected for interviews regarding certain error codes were asked the same Items during the interview, and they were asked to explain how they solved all the Items on the test. During this process, findings were obtained showing that the teacher candidates selected for the interview also exhibited different error codes than those intended for their selection. Consistent with this interpretation, when asked to re-solve Item 1 during the interview, Tc-4 and Tc-7 reached the correct answer. In contrast, Tc-12's explanations included shifting referents (e.g., moving between “100 g water,” “100 g solution,” and “100 mL solution”) and a

density-based framing of concentration. Tc-12 repeated the same choice upon re-solving and continued to shift between “water” and “solution” referents, suggesting a less stable solvent-solution boundary in this individual case. Therefore, DC-1 selections are interpreted cautiously: although the distractor targets referent confusion, interview evidence indicates that at least some DC-1 selections reflect rapid responding without verification. Because DC-1 is captured by a single item, we avoid person-level claims and treat interview evidence as analytic exemplars that qualify the intended diagnostic interpretation.

DC-2 error – difficulty establishing the solvent volume-solution volume distinction. The DC-2 error was identified specifically in Items 5 and 9. In the interview with Tc-2 who was selected for this error code, the teacher candidate defined molarity as “the amount of substance dissolved in one liter of water in moles” and molality as “the number of moles dissolved in one kilogram of solution,” overlooking the critical solvent-solution distinction. The explanation suggests that the decision was guided primarily by unit symbols (mol L^{-1} vs. mol kg^{-1}) rather than by explicitly checking what the unit is “per” (solution volume vs. solvent mass). This DC-2-targeted choice co-occurred with cue-based unit reliance (SD-1) and/or formula-level confusion (UM-2), as summarized in Appendix A2. Similar “unit-symbol first” routes were evident in other interviews: for example, when she re-solved Item 5, Tc-10 treated V as solvent volume (“ V refers to the solvent”), and Tc-9 reported difficulty deciding whether to use solution volume or solvent volume in molarity contexts. At the same time, Tc-8 initially overlooked the “solution” referent in Item 5 (“I only took the water”) but corrected the item upon re-solving, illustrating that some DC-2-targeted selections can also reflect unstable referent checking rather than fixed beliefs.

UM-1 error – misinterpretation of unit representation. UM-1 was observed particularly in Item 1 through Option B. In the interview with Tc-3, although the teacher candidate correctly defined mass percentage concentration, the choice was justified as follows: “It says the solution is 100 millilitres and there are 20 grams of dissolved salt. So, it matches my definition.” For this teacher candidate, mL and g were treated as functionally interchangeable in the item context; when asked directly whether millilitres and grams are the same, the teacher candidate confirmed “Yes.” Thus, the difficulty appears at the transfer point from a recalled definition to contextualised unit reasoning, rather than at the level of definitional recall itself. This case is reported as an exemplar of UM-1 reasoning to document a plausible mechanism, not as evidence that all teacher candidates equate these units.

UM-2 error – confusion of molarity and molality formulas. UM-2 was observed particularly in Items 5 and 9. In the interview with Tc-4, molality was expressed as “ $M = n/V$, mol L^{-1} ,” and the teacher candidate reported remembering the concept based on the letter “l” in “molality,” interpreted as “liter.” This indicates reliance on a superficial linguistic cue rather than on the definitional referent, and it illustrates that the UM-2 target can be realized through a cue-based (SD-1)



route rather than only through a stable definitional misconception. Related phrasing appeared in some other interviews as well (*e.g.*, defining molality as “moles per kilogram of solution” or describing the distinction as “one is liter, one is kilogram” without explicitly stating “per liter of solution” *versus* “per kilogram of solvent”), suggesting that surface-level unit cues can dominate over referent-level checks in molarity–molality distinctions.

RR-1 error – failure to establish a proportional relationship.

RR-1 was observed particularly in Items 2, 6, and 10. In the interview with Tc-5, the teacher candidate initially selected an incorrect option but, during re-solving of Item 10, articulated the correct proportional reasoning: “Since it is mol per kilogram, the mass of the solvent is reduced due to evaporation... the denominator is reduced, so the concentration increases.” Interview data therefore suggest that the proportional schema was available but was not consistently activated under the monitoring conditions of the test. Similarly, Tc-7 produced the correct solution when re-solving Item 6 during the interview (“the amount of solute increased... the volume is constant... concentration should increase”) yet reported marking a different option during the test (“Maybe I didn’t read it carefully”), indicating variability in attention and checking across contexts.

RR-2 error – incorrect establishment of a proportional relationship. RR-2 was observed in Items 2, 6, and 10. When Tc-6 re-solved Item 6, the teacher candidate used the formula correctly and reached the correct conclusion. However, the test choice was justified by an overgeneralized heuristic: “I marked it thinking it wouldn’t change if there was no change in volume and moved on.” This indicates that a single cue (“volume does not change”) can override context-specific checks during the test. Tc-6 also described missing a key conversion in a molality context (“I probably missed that 1000 g is 1 kg”), illustrating that conversion/monitoring slips can co-occur with overgeneralized proportional rules.

SD-1 error – superficial decision (focusing on a single cue).

SD-1 was identified through Option C selections in Items 3, 7, and 11, which were designed to attract responses based on a salient cue (*e.g.*, an apparently simple numeric relation) without coordinating all required task elements. Across interviews, teacher candidates often reported relying on a single numerical cue without checking whether unit conversions or referents were required. For example, when Tc-10 explained how she solved Item 11, she stated, “I didn’t convert to kilograms because they were all in grams... In C, $30/40 \approx 3/4$,” indicating use of a convenient numeric comparison rather than verification of unit meaning and conversion steps. Notably, this shortcut was applied again during re-solving and led to the same incorrect choice, supporting SD-1 as a relatively stable decision strategy for that teacher candidate in this task context.

Although Tc-9 was initially selected due to a response pattern mapped to CM-1, the interview explanation did not include any attempted calculation or conversion. Instead, the choice was justified through a salient linguistic association (*e.g.*, linking molality to “liter”), which is discussed here as cue-

based responding (SD-1) accompanied by molarity–molality unit confusion (UM-2). This illustrates that the option–code mapping captures diagnostic intent, whereas interviews clarify the reasoning route underlying a selection.

CM-1 error – execution breakdown with loss of unit-referent meaning during processing. CM-1 was observed in Items 3, 4, 7, 8, 11, and 12. Quantitative analysis showed that 59 teacher candidates exhibited this pattern in mass-percent items, 27 in molarity items, and 62 in molality items. Interview data suggest that CM-1 often reflects “knowing the information but not consistently activating it during processing.” Tc-8 reported being able to define the molarity, molality, and mass percent concepts correctly, yet selected incorrect options during the test and could not fully explain why; during re-solving Item 1, Tc-8 revised after reconsidering the referent (“I only took the water”) and corrected the item, consistent with a process-level slip rather than lack of declarative knowledge. Similarly, Tc-9 reported remembering the molarity formula but confusing solvent and solution volumes; although the item was corrected during re-solving, the teacher candidate described making a quick decision (“I may have been confused when doing the calculation quickly”). Taken together, CM-1-coded selections should be interpreted cautiously, because interviews indicate that some cases may be better described as cue-based responding and/or unit confusion routes rather than loss of unit-referent tracking or procedural monitoring alone (see Appendix A2).

Across interviews, several teacher candidates demonstrated that a given distractor selection can reflect momentary monitoring lapses, cue-based associations, or conversion omissions rather than a stable misconception. Conversely, some teacher candidates showed correct procedural performance while still expressing definitional ambiguity, indicating that response patterns are best interpreted as item-level evidence within the instrument’s formats. Taken together, the interview findings support interpreting error codes as item-level diagnostic signals within this instrument, while recognizing that the same signal may reflect different underlying routes across individuals and occasions.

Discussion

This study aimed to examine how teacher candidates reason about solution concentration units and to interpret recurring error patterns through an integrated theoretical framework. The following discussion briefly situates these findings within the relationships among the theoretical perspectives guiding this study.

The integrative explanatory model developed to illustrate the described relationships between the 4 theories is shown in Fig. 2. While CLT, DPT, RCT, and CCT are established perspectives in the literature, the relations shown in the figure represent the authors’ synthesis of how these perspectives explain the error patterns identified in the test and interview data. Taken together, the findings suggest that these errors are better understood not as isolated mistakes, but as recurring



difficulties in coordinating unit meaning, referent, representation, and operation across tasks.

From a CLT perspective, items that require the coordination of solute–solvent distinctions, unit referents, conversions, and proportional reasoning can increase task demands, making sustained representational coordination (RCT) and analytic checking (Type 2; DPT) less likely in the moment and thereby constraining the conditions that support conceptual restructuring (CCT). Conversely, representational supports (RCT) and clearer conceptual boundaries (CCT) can help reduce demands by making key distinctions more explicit and stabilising meaning across contexts (*i.e.*, supporting boundary stability for molarity–molality and solution–solvent referents across item contexts). The DPT–CCT link captures that analytic checking can support restructuring, while more coherent conceptions can facilitate engagement of analytic checking during problem solving. The RCT–CCT link reflects that representational coordination can surface inconsistencies that prompt restructuring and that coherent conceptions enable more stable symbol–referent mapping and thus help maintain boundary stability when learners move between definitional statements and contextual application items. The DPT–RCT link is conditional: salient representational cues may bias Type 1 responding, whereas prompts that make mismatches explicit (*e.g.*, unit-referent cues in item wording) may facilitate Type 2 checking. The CLT–CCT link denotes a theoretically grounded coupling between task demands and the conditions for restructuring rather than implying immediate conceptual change. Within this framework, metacognitive monitoring and regulation are treated not as an additional theory, but as a cross-cutting self-regulatory process that supports conflict detection, checking, and coordination across formula–unit–context during concentration-related reasoning (Flavell, 1979; Schraw and Dennison, 1994). Overall, this framework provides a strong analytical basis for classifying and interpreting pre-service teachers' definitional and procedural error patterns within the evidential scope of the diagnostic items and interview justifications.

The findings of this study suggest that teacher candidates experience multidimensional difficulties in understanding solution concentration units (mass percent, molarity, and molality), especially when tasks require coordinating unit meaning with the correct referent and carrying out multi-step operations. Across the diagnostic test and interviews, these difficulties appeared not only as isolated errors but also as recurring reasoning patterns shaping how candidates interpreted and coordinated chemical representations. These patterns are consistent with prior work showing that learners often struggle to connect symbolic expressions and unit labels to their intended meanings and to coordinate different representations during problem solving (Johnstone, 1991, 2006; Kozma and Russell, 2005; Wu, 2003).

From a CCT perspective, this suggests that the molarity–molality and solution–solvent boundaries may not be fully stable across contexts, which can lead to shifting meanings between definitional statements and contextualised items.

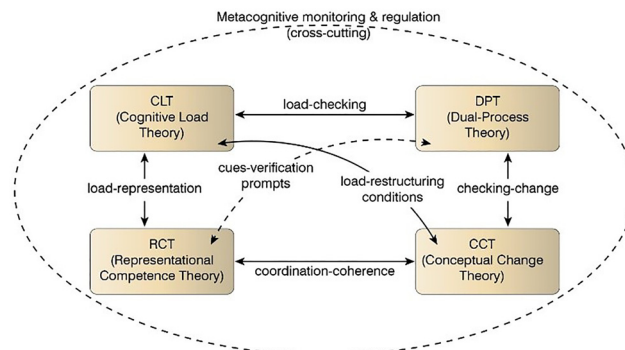


Fig. 2 Authors' integrative model linking CLT, DPT, CCT, and RCT in the interpretation of recurring concentration-related error patterns. The diagram summarises reciprocal relations among cognitive load (CLT), processing mode (DPT), conceptual restructuring (CCT), and representational coordination (RCT). Arrow labels indicate the focal mechanism linking each pair: load-checking (CLT ↔ DPT), load-representation (CLT ↔ RCT), checking-change (DPT ↔ CCT), coordination-coherence (RCT ↔ CCT), and load-restructuring conditions (CLT ↔ CCT). The dashed ellipse denotes metacognitive monitoring and regulation as a cross-cutting process supporting conflict detection, checking, and coordination across formula–unit–context. Solid arrows denote robust reciprocal couplings among constructs; dashed links denote conditional relations. The dashed DPT–RCT link (cues-verification prompts) is explicitly conditional: salient cues may bias Type 1 responding, whereas prompts that surface mismatches may facilitate Type 2 checking.

From an RCT perspective, the results are consistent with difficulties linking unit symbols to their intended referents, such as treating “L” as solvent volume rather than solution volume in molarity or not consistently treating “kg” as solvent mass in molality. At the process level, DPT is consistent with interview accounts showing that some teacher candidates relied on a salient cue rather than explicitly checking what the unit was “per.”

CLT is used here as a task-demand lens. Items requiring coordination of unit meaning, referent, conversions, and proportional reasoning were more likely to create difficulty. Because cognitive load was not measured directly, these claims are limited to interpretations of item demands and observed response patterns. Other influences, such as prior instruction, time pressure, or anxiety, may also have contributed to quick responding and reduced checking, but these are treated here as plausible influences rather than tested explanations.

Overall accuracy was relatively high, but performance varied across item types and demands. Performance tended to be stronger on items requiring more direct descriptive or qualitative reasoning, and weaker on items requiring stable coordination of unit meaning, referents, conversions, and multi-step proportional reasoning, particularly in the molarity and molality contexts. These patterns suggest that many teacher candidates could recall key terms and symbols or apply familiar procedures, but did not always connect that knowledge to problem situations requiring coordinated referent checking and proportional reasoning (Johnstone, 1991, 2006; Kozma and Russell, 2005; Raviolo *et al.*, 2021; Stott, 2023).



Interview accounts helped clarify the reasoning routes behind the lowest-performing items (Items 5, 9, and 11). For Item 5, the low correct-response rate suggests that applying molarity in context was challenging for a subset of teacher candidates; several responses reflected attention to solvent volume rather than total solution volume, indicating a weak linkage between the formal definition and contextual application. Interview explanations further suggested that some teacher candidates relied on symbolic cues (e.g., “mol L⁻¹”) and proceeded quickly, with limited analytic checking in the moment. For Item 9, performance patterns suggest that distinguishing molarity from molality also posed difficulties: some teacher candidates used unit symbols (e.g., “mol L⁻¹”, “mol kg⁻¹”) as cues but did not consistently connect these cues to the underlying referent distinction in their explanations. For Item 11, several responses indicated difficulty coordinating proportional reasoning and unit conversion simultaneously. For instance, some teacher candidates approximated ratios in a visually convenient form (e.g., 30/40 \approx 3/4) while overlooking the gram-kilogram conversion, leading to plausible but contextually incorrect conclusions. Taken together, these cases highlight difficulties in integrating unit meaning, unit conversion, and proportional reasoning within the scope of the instrument.

In addition, interviews indicated that some teacher candidates drew on mental imagery and everyday contexts when making sense of concentration. For example, one participant (Tc-7) described concentration increase using a particle-based visualization (e.g., fewer “empty circles” and more “full circles” after evaporation), suggesting an inclination toward representational thinking, albeit with limited scientific precision in parts of the explanation. Similarly, some teacher candidates (e.g., Tc-11) used everyday analogies such as diluted tea colour or adding water to fruit juice to articulate concentration change. Such analogies may provide an initial entry point for reasoning; however, the interview data suggest that everyday associations were not always successfully coordinated with formal representational and quantitative reasoning. When asked how to prepare a one-molar sugar solution, several teacher candidates provided incomplete procedural accounts or required additional prompts, suggesting difficulties in connecting laboratory-oriented procedures with unit definitions. More broadly, these findings suggest that some teacher candidates

experienced difficulty in coordinating solute–solvent–solution distinctions, linking symbolic expressions to contextual meaning, and integrating proportional reasoning with unit conversion.

To move beyond overall accuracy patterns, we also examined whether teacher candidates’ definitional choices and performance on application items differed across concentration units. Pairwise comparisons (see Table 6) showed unit-dependent definition–application patterns. For mass percent, definitional-statement selection and accuracy on application items were similar (80.3% vs. 76.3%). For molarity and molality, application-item accuracy was substantially higher than definitional-statement selection (M: 32.3% vs. 84.8%; m: 44.7% vs. 84.8%). Because the definitional items asked teacher candidates to identify the correct definition among alternative statements, these results should be interpreted as a format-bounded comparison within this instrument. Within this scope, the pattern suggests a definition–application mismatch for molarity and molality: some teacher candidates selected correct options in application items while not consistently selecting definitional statements that explicitly fit the unit referent (molarity = per liter of solution; molality = per kilogram of solvent). Interview justifications added task-level detail. Several teacher candidates described cue-based recall (e.g., “I remembered the formula,” “the word liter came to mind”) and reported limited verification during the test, suggesting that correct responding in operational items can be achieved *via* rapid retrieval and execution without explicit referent checking.

In contrast, other teacher candidates described an explicit unit-referent checking routine by stating what the unit was “per,” identifying which quantity in the prompt corresponded to that referent, and noting required conversions (mL \rightarrow L; g \rightarrow kg), especially when re-solving items during interviews. In the coding framework, cue-based routes are consistent with SD-1-type responding, whereas difficulties coordinating unit symbols with their intended referents are reflected in item-level patterns such as DC-2 and UM-2. Importantly, these codes capture the diagnostic intent of distractors and the reasoning routes observed in specific tasks rather than stable person-level misconceptions.

Instructional implications

The implications below are offered as theory-informed possibilities derived from the response patterns and interview

Table 6 Definition–application alignment across concentration units

Concentration unit	Descriptive accuracy (%)	Operational accuracy (%)	Δ (operational–descriptive)	Comment
Mass percentage (%w/w)	80.3	76.3	−4.0	Close alignment within the instrument formats; definitional selection and application performance are broadly similar.
Molarity (M)	32.3	84.8	+52.5	Large mismatch; high operational accuracy co-occurs with low selection of definitional statements that fix the referent (“per liter of solution”).
Molality (m)	44.7	84.8	+40.1	Substantial mismatch; operational success can occur without consistent selection of definitional statements that fix the referent (“per kilogram of solvent”).

Note. Definitional items (1, 5, 9) required selection of provided definitional statements (not constructed explanations). Application items (4, 8, 12) required selecting the correct option in an algorithmic context. Δ is calculated as (operational % – definitional %).



accounts in this study; they are not presented as evidence about the effectiveness of particular instructional interventions.

First, difficulties distinguishing solution–solvent referents and differentiating molarity from molality suggest that definitional knowledge may be recalled verbally yet not consistently anchored to the correct referent during application items. For example, low performance on Items 5 and 9 indicates that some teacher candidates did not reliably coordinate unit meaning with its intended referent.

Accordingly, instructors and instructional designers may consider supports that make unit meaning and its referent explicit during problem solving. Prior research also suggests that representational supports can improve problem solving when they make symbol-referent relationships more explicit. For example, Ralph and Lewis (2020) found that assessment formats incorporating structured representations, such as tables that helped students recognize how the same unit could apply across different chemical contexts, improved student performance. A brief referent-check routine can therefore be embedded before and after calculations: learners write what the unit is “per” (*e.g.*, per litre of solution; per kilogram of solvent), identify which quantity in the prompt matches that referent, and then verify that the denominator used in the calculation matches the referent in context. Such routines may support more consistent symbol-referent mapping (RCT) and may help learners maintain key distinctions (*e.g.*, solution *vs.* solvent; molarity *vs.* molality) across contexts (CCT).

Interview data also suggest the potential value of variation tasks that make variable roles explicit (*e.g.*, preparing parallel examples using % w/w, M, and m to highlight what is held constant and what changes). Representation-matching activities may further help learners connect symbolic expressions (% w/w, mol L⁻¹, mol kg⁻¹) to the quantities they refer to in context.

In addition, several interview accounts reflected rapid responding, reliance on a single salient cue, and limited monitoring. Instructional routines such as “initial choice + short justification,” think-aloud or error-analysis tasks, and step-by-step prompts (*e.g.*, “What does this step represent?”) may encourage checking and monitoring during multi-step concentration problems.

At the program level, teacher education may benefit from modules that connect common concentration-unit difficulties to practical supports aligned with CLT, DPT, CCT, and RCT, thereby strengthening candidates’ own reasoning and their capacity to anticipate learners’ difficulties. Table 7 summarises example strategies related to the cognitive processes suggested by the present diagnostic patterns.

Conclusion

A key contribution of the study is the use of a theory-informed diagnostic framework (CLT, DPT, CCT, and RCT) to interpret not only error frequencies but also the item-level reasoning routes underlying selected distractors. Rather than claiming to measure these theories directly, the study shows how they can be used together to explain common concentration-unit errors as the result of interacting task demands, processing tendencies, conceptual boundary instability, and representational coordination.

Overall, the results suggest that supporting learners’ reasoning about concentration may require more than procedural practice. Learners may benefit from routines and tasks that explicitly connect unit meaning to its referent, strengthen coordination across representations, and support monitoring during problem solving.

Limitations and future research

This study has certain limitations. First, the study was conducted with 152 teacher candidates from a single context; including teacher candidates from different universities and disciplines would strengthen the generalizability of the findings.

Second, the diagnostic instrument consisted of 12 multiple-choice items and relied on predefined, theory-informed error categories. Accordingly, teacher candidates’ ways of making sense of concentration concepts were interpreted within the response formats and diagnostic scope of this instrument. Future research could incorporate open-ended questions, performance-based assessments, and/or constructed-response

Table 7 Theory-based instructional strategies and expected outcomes

Theory	Targeted cognitive process	Teaching strategy/example activity	Expected outcomes
CCT	Conceptual restructuring	Simultaneous preparation of three concentration units (%w/w, M, m)	Concretizing the solvent–solute distinction and the variables in the definition
RCT	Representational coordination	Unit–referent mapping and representation-matching tasks (symbolic–numerical–contextual)	Stronger symbol–referent integrity; reduced unit/referent confusions
CLT	Managing task demands	Step-by-step problem solving, asking “What does each step represent?”	Reduced execution slips; improved coordination under multi-step demand
DPT	Balancing intuitive <i>vs.</i> analytic processing	“Quick choice + justification”/cognitive deceleration tasks	Increased checking; reduced single-cue responding
Cross-cutting	Metacognitive monitoring	Error analysis, think-aloud, and self-reflection activities	Improved awareness of reasoning routes; more consistent verification

Note. Each strategy targets cognitive challenges suggested by the diagnostic patterns and interview-derived reasoning routes; implications are presented as theory-informed recommendations bounded to the formats and evidence of this study.



explanations to capture reasoning routes that may not surface in option selections.

Third, because the test did not include external visual representations (*e.g.*, particulate diagrams, graphs, or dynamic simulations), RCT interpretations in this study are limited to symbolic-numerical-contextual coordination within text-based items. Future work should extend the instrument to include additional representational formats to examine whether similar coordination difficulties emerge across visual and particulate-level representations.

Fourth, data collection was limited to the diagnostic test and semi-structured interviews. Interview prompts were anchored to teacher candidates' item choices, which strengthens triangulation for those items but also bounds the analysis to patterns that are detectable through this design. Process-based methods (*e.g.*, classroom observation, video-based problem-solving sessions, eye-tracking, or digital learning analytics) could be used in future research to capture decision points, checking behaviours, and inter-representational transitions more dynamically, thereby providing finer-grained evidence about metacognitive monitoring and reasoning control.

Fifth, although no grade-level comparisons were conducted in the present study, future research could incorporate grade-level analyses to examine in more detail how definitional and procedural difficulties may vary across stages of teacher education.

Finally, because the study examined only the reasoning patterns associated with incorrect options, it cannot determine whether some correct responses were also reached through surface cues or intuitive shortcuts. Correct responses may not always reflect coherent reasoning pathways.

Author contributions

N. G. developed the theoretical framework and designed the research. Both authors conducted the data collection process and jointly validated all data. A. R. E. conducted the interviews, performed the data analysis, and prepared the graphical visualizations. N. G. carried out the theoretical interpretation of the diagnostic test findings. Both authors contributed to writing the initial draft and collaboratively reviewed and approved the final version of the manuscript.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

The data supporting this study were obtained from human participants and are not publicly available due to privacy and ethical considerations. Data may be made available from the corresponding author upon reasonable request and with approval from the Ethics Committee of Eskişehir Osmangazi University (Decision No: 2025-14).

Appendices

Appendix A1. Diagnostic test

Item no.	Item	Options
1	Which of the following solutions has a mass percent concentration of 20%?	(A) A solution prepared by dissolving 20 grams of table salt in 100 grams of water (B) A solution containing 20 grams of dissolved salt in 100 milliliters of table salt solution (C) A solution prepared by dissolving 20 grams of table salt in 80 grams of water
2	Pure water is added to a 20% KNO ₃ solution at a constant temperature. How does the mass percentage concentration of the solution change?	(A) Increases (B) Decreases (C) Remains unchanged
3	Which of the following solutions has a higher mass percentage concentration?	(A) A mixture of 20 grams of sugar and 30 grams of water (B) A mixture of 30 grams of sugar and 40 grams of water (C) A mixture of 50 grams of sugar and 80 grams of water
4	25 grams of sugar is dissolved in 100 grams of sugar solution containing 25% by mass. What is the percentage by mass of the new solution?	(A) 50% (B) 40% (C) 30%
5	Which of the following solutions has a concentration of 1 molar?	(A) Solution obtained by dissolving 1 mole of sugar in 1 L of water (B) Solution obtained by dissolving 1 mole of sugar in 1 kg of water (C) 1 L of solution obtained by dissolving 1 mole of sugar in water
6	A quantity of solid NaOH is added to an unsaturated 0.5 M NaOH solution and allowed to dissolve. How does the molar concentration of the solution change? (It will be assumed that there is no increase in volume.)	(A) Increases (B) Decreases (C) Remains unchanged
7	Which of the following solutions contains a larger amount of solute?	(A) 1 molar 400 mL NaOH solution (B) 2 molar 300 mL NaOH solution (C) 3 molar 100 mL NaOH solution
8	How many moles of dissolved KOH are there in 500 mL of a 0.4 M KOH solution?	(A) 0.8 (B) 0.4 (C) 0.2
9	Which of the following solutions has a concentration of 1 molal?	(A) A 1 L solution prepared by dissolving 1 mole of table salt in water (B) A solution prepared by dissolving 1 mole of table salt in 1000 g of water (C) A 1 kg solution prepared by dissolving 1 mole of table salt in water



Table (continued)

Item no.	Item	Options
10	A certain amount of water evaporates from an unsaturated 0.2 molal KNO_3 solution without precipitation. How does the molal concentration of the solution change?	(A) Increases (B) Decreases (C) Remains unchanged
11	Which of the following solutions has a higher molal concentration? (NaOH : 40 g mol^{-1} , KOH : 56 g mol^{-1})	(A) Solution prepared by dissolving 20 grams of solid NaOH in 500 grams of water (B) Solution prepared by dissolving 24 grams of solid KOH in 500 grams of water (C) Solution prepared by dissolving 30 grams of solid NaOH in 1000 grams of water
12	How many moles of dissolved NaOH are there in a 0.4 molal NaOH solution prepared with 500 grams of water?	(A) 0.8 (B) 0.4 (C) 0.2

Appendix A2. Interview-test map (test selection → intended code target → interview-derived reasoning route)

Teacher candidate (Tc-#)	Test selection pattern (item-option)	Intended code target (from option-code map)	Interview-derived reasoning route (as observed)	Illustrative excerpt (translated)	Re-solve outcome
Tc-1	Item 1-A (mass %); Item 9-C (molality)	DC-1; DC-2	Referent shift + limited checking: could state definitions when probed, yet framed molality as kg of solution rather than kg of solvent; also reported quick marking without verification.	“40 g NaOH + 960 g water ... equals 1 molal,” indicating kg-of-solution framing. “I may have marked it directly and moved on.”	Item 1 corrected, Item 9 not corrected
Tc-2	Item 5-A (molarity); Item 9-C (molality)	DC-2	Unit-symbol first (cue-based) + referent-check failure: relied on unit symbols (mol L^{-1} vs. mol kg^{-1}) and described molarity as per liter of water and molality as per kg of solution, without explicitly checking solvent vs. solution referents.	“Molarity is ... in 1 L of water ...; molality is ... in 1 kg of solution.” “One is in liters; the other is in kilograms.” (but described molality as per kg of solution)	Item 5 and 9 not corrected
Tc-3	Item 1-B (mass %)	UM-1	Unit interchangeability assumption: matched numeric values across units and treated mL and gram interchangeable; applied a recalled definition without unit-consistency checking.	“It says 100 mL solution and 20 g solute ... so it matches my definition.”	Item 1 not corrected
Tc-4	Item 5-B (molarity); Item 9-A (molality)	UM-2	Cue-based mnemonic (linguistic cue): used a superficial mnemonic (“l in molality → liter”) and linked concept names to unit symbols, producing systematic formula/concept mismatches.	“‘l’ in molality ... the l of liter (volume),” guiding the selection.	Item 5 and 9 not corrected
Tc-5	Item 9-C (molality), Item 10-B (molality)	RR-1	Proportional schema available; test slip/monitoring lapse: demonstrated correct proportional reasoning during re-solving; the test error was explained as fast responding rather than lack of proportional reasoning. Also showed occasional definitional referent slippage in verbal explanations.	“Since the denominator decreases (evaporation), the molal concentration increases.”	Item 9 and 10 corrected
Tc-6	Item 6-C (molarity); Item 9-C (molality) Item 10-C (molality);	RR-2; DC-2	Overgeneralised heuristic + conversion omission: used a single cue (“volume constant → no change”) and missed a key conversion ($1000 \text{ g} = 1 \text{ kg}$) during fast responding; corrected when prompted to re-solve.	“I probably took ‘volume is constant’ ... so I thought it wouldn’t change.”	Item 6,9, and 10 corrected



Table (continued)

Teacher candidate (Tc-#)	Test selection pattern (item-option)	Intended code target (from option-code map)	Interview-derived reasoning route (as observed)	Illustrative excerpt (translated)	Re-solve outcome
Tc-7	Item 7-C (Molarity); Item 11-C (molality)	SD-1	Single-cue responding + limited reading/verification: reported not reading carefully and relied on a salient cue without coordinating all required steps (referent/conversion/proportion).	"I guess I didn't read carefully... volume is constant."	NR
Tc-8	Item 3-A (mass %); Item 4-A (mass %)	CM-1	Execution breakdown/monitoring lapse: could produce a relevant ratio yet still selected an incorrect option, suggesting loss of monitoring/definitional focus during procedure.	"I wrote 20/120... but I still marked A; I don't know why."	Item 3 and 4 corrected
Tc-9	Item 3-A (mass %); Item 7-A (molarity); Item 9-C (molality)	CM-1; DC-2	Comparison uncertainty + referent confusion; improved with prompting: computed ratios but struggled to compare them confidently; also mixed solvent/solution referents in definitional contexts; performance improved when asked to re-solve.	"I found 20/50 and 30/200, but I couldn't compare them confidently."	Item 3 corrected (partial), Item 7 corrected, Item 9 not corrected
Tc-10	Item 5-A (Molarity); Item 6-C (molarity); Item 7-A (molarity)	DC-2; RR-2; CM-1	Mixed route (referent confusion + cue-based responding + occasional arithmetic slips): expressed low certainty and incomplete checking; combined solvent-solution confusion with quick responding and occasional procedural slips.	"I'm never 100% sure of myself," indicating limited verification during solving.	Item 5 not corrected, Item 6 and 7 corrected
Tc-11	Item 9-A (molality), Item 10-C (molality), Item 11-B (molality), Item 12-B (molality)	UM-2; RR-1; CM-1	Low familiarity/guessing + cue-based name-unit associations: reported guessing; selections reflected surface associations rather than definitional referent checking (note that some statements still show solvent-focused framing).	"I didn't convert... In C, 30/40 \approx 3/4." "I have no idea about molality... I answered randomly." "V is the solvent's volume... we only take the solvent."	NR
Tc-12	Item 1-A (mass %); Item 5-A (Molarity); Item 9-C (molality)	DC-1; DC-2 (plus mixed profile)	Mixed profile; referent shifts + fast responding: interview suggests that some selections reflect quick marking and shifting referents rather than stable misconceptions; repeatedly treated "kg" as "kg of solution" and defined concentration <i>via</i> density.	"I first marked A on Item 1... and chose C on Item 9," later revising reasoning when probed. Defined concentration <i>via</i> density; repeatedly treated "kg" as "kg of solution."	Item 1 not corrected, Item 5 NR, Item 9 not corrected

Note. This table increases transparency regarding (a) how interview teacher candidates were selected based on diagnostic-test response patterns and (b) how interview explanations clarify the reasoning routes behind those selections. Error codes represent the diagnostic intent of distractors (option-code map), not stable person-level misconceptions. *Excerpts were translated into English by the authors and lightly edited for clarity without changing meaning.* Re-solve outcome" indicates whether the teacher candidate corrected the relevant item(s) when asked to re-solve during the interview (corrected/not corrected/NR = not reported).

References

- Cohen L., Manion L. and Morrison K., (2018), *Research methods in education*, 8th edn., Routledge, DOI: [10.4324/9781315456539](https://doi.org/10.4324/9781315456539).
- Demircioğlu G., Ayas A. and Demircioğlu H., (2005), Conceptual change achieved through a new teaching program on acids and bases, *Chem. Educ. Res. Pract.*, **6**(1), 36–51, DOI: [10.1039/B4RP90003K](https://doi.org/10.1039/B4RP90003K).
- Denzin N. K., (1978), *The research act: a theoretical introduction to sociological methods*, 2nd edn, McGraw-Hill.
- Duit R. and Treagust D. F., (2003), Conceptual change: a powerful framework for improving science teaching and learning, *Int. J. Sci. Educ.*, **25**(6), 671–688, DOI: [10.1080/09500690305016](https://doi.org/10.1080/09500690305016).
- Evans J. S. B. T., (2008), Dual-processing accounts of reasoning, judgment, and social cognition, *Annu. Rev. Psychol.*, **59**(1), 255–278, DOI: [10.1146/annurev.psych.59.103006.093629](https://doi.org/10.1146/annurev.psych.59.103006.093629).



- Flavell J. H., (1979), Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry, *Am. Psychol.*, **34**(10), 906–911, DOI: [10.1037/0003-066X.34.10.906](https://doi.org/10.1037/0003-066X.34.10.906).
- Gabel D., (1999), Improving teaching and learning through chemistry education research: a look to the future, *J. Chem. Educ.*, **76**(4), 548–554, DOI: [10.1021/ed076p548](https://doi.org/10.1021/ed076p548).
- Gilbert J. K. and Justi R., (2016), *Models and modeling in science education*, Springer, DOI: [10.1007/978-3-319-29039-3](https://doi.org/10.1007/978-3-319-29039-3).
- Johnstone A. H., (1991), Why is science difficult to learn? Things are seldom what they seem, *J. Comput. Assist. Learn.*, **7**(2), 75–83, DOI: [10.1111/j.1365-2729.1991.tb00230.x](https://doi.org/10.1111/j.1365-2729.1991.tb00230.x).
- Johnstone A. H., (1993), The development of chemistry teaching: a changing response to changing demand, *J. Chem. Educ.*, **70**(9), 701–705, DOI: [10.1021/ed070p701](https://doi.org/10.1021/ed070p701).
- Johnstone A. H., (2006), Chemical education research in Glasgow in perspective, *Chem. Educ. Res. Pract.*, **7**(2), 49–63, DOI: [10.1039/B5RP90021B](https://doi.org/10.1039/B5RP90021B).
- Kahneman D., (2011), *Thinking, fast and slow*, Farrar, Straus and Giroux.
- Kozma R. B. and Russell J., (2005), Teacher candidates becoming chemists: developing representational competence, in Gilbert J. K. (ed.), *Visualization in science education*, Springer, pp. 121–145, DOI: [10.1007/1-4020-3613-2_8](https://doi.org/10.1007/1-4020-3613-2_8).
- Merriam S. B., (1998), *Qualitative research and case study applications in education*, Jossey-Bass.
- Naah B. M. and Sanger M. J., (2012), Student misconceptions in writing balanced equations for dissolving ionic compounds in water, *Chem. Educ. Res. Pract.*, **13**(3), 186–194, DOI: [10.1039/C2RP00015F](https://doi.org/10.1039/C2RP00015F).
- Nakhleh M. B., (1992), Why some teacher candidates don't learn chemistry: chemical misconceptions, *J. Chem. Educ.*, **69**(3), 191, DOI: [10.1021/ed069p191](https://doi.org/10.1021/ed069p191).
- Patton M. Q., (1999), Enhancing the quality and credibility of qualitative analysis, *Health Services Res.*, **34**(5 Pt 2), 1189–1208.
- Pınarbaşı T. and Canpolat N., (2003), Teacher candidates' understanding of solution chemistry concepts, *J. Chem. Educ.*, **80**(11), 1328–1332, DOI: [10.1021/ed080p1328](https://doi.org/10.1021/ed080p1328).
- Posner G. J., Strike K. A., Hewson P. W. and Gertzog W. A., (1982), Accommodation of a scientific conception: toward a theory of conceptual change, *Sci. Educ.*, **66**(2), 211–227, DOI: [10.1002/sce.3730660207](https://doi.org/10.1002/sce.3730660207).
- Ralph V. R. and Lewis S. E., (2020), Impact of representations in assessments on student performance and equity, *J. Chem. Educ.*, **97**(3), 603–615, DOI: [10.1021/acs.jchemed.9b01058](https://doi.org/10.1021/acs.jchemed.9b01058).
- Raviolo A., Farré A. S. and Schroh N. T., (2021), Teacher candidates' understanding of molar concentration, *Chem. Educ. Res. Pract.*, **22**, 486–497, DOI: [10.1039/d0rp00344a](https://doi.org/10.1039/d0rp00344a).
- Schraw G. and Dennison R. S., (1994), Assessing metacognitive awareness, *Contemp. Educ. Psychol.*, **19**, 460–475, DOI: [10.1006/ceps.1994.1033](https://doi.org/10.1006/ceps.1994.1033).
- Sheppard K., (2006), High school teacher candidates' understanding of titrations and related acid–base phenomena, *Chem. Educ. Res. Pract.*, **7**(1), 32–45, DOI: [10.1039/B5RP90014J](https://doi.org/10.1039/B5RP90014J).
- Staver J. R. and Jacks T., (1988), The influence of cognitive reasoning level, cognitive restructuring ability, disembedding ability, working memory capacity, and prior achievement on teacher candidates' performance on balancing equations by inspection, *J. Res. Sci. Teach.*, **25**(9), 763–775, DOI: [10.1002/tea.3660250906](https://doi.org/10.1002/tea.3660250906).
- Stott A. E., (2023), The efficacy of instruction in application of mole ratios and submicro- and macroscopic equivalent forms of the mole within the unit factor method, *Chem. Educ. Res. Pract.*, **24**, 551–566, DOI: [10.1039/D2RP00245K](https://doi.org/10.1039/D2RP00245K).
- Sweller J., (1988), Cognitive load during problem solving: effects on learning, *Cogn. Sci.*, **12**(2), 257–285, DOI: [10.1207/s15516709cog1202_4](https://doi.org/10.1207/s15516709cog1202_4).
- Sweller J., (1994), Cognitive load theory, learning difficulty, and instructional design, *Learn. Instruct.*, **4**(4), 295–312, DOI: [10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5).
- Sweller J., (2010), Element interactivity and intrinsic, extraneous, and germane cognitive load, *Educ. Psychol. Rev.*, **22**(2), 123–138, DOI: [10.1007/s10648-010-9128-5](https://doi.org/10.1007/s10648-010-9128-5).
- Taber K. S., (2002), *Chemical misconceptions-prevention, diagnosis and cure: theoretical background*, vol. 1, London: Royal Society of Chemistry.
- Talanquer V., (2009), On cognitive constraints and learning progressions: the case of structure of matter, *Int. J. Sci. Educ.*, **31**(15), 2123–2136.
- Talanquer V., (2014), Chemistry education: ten heuristics to tame, *J. Chem. Educ.*, **91**(8), 1091–1097, DOI: [10.1021/ed4008765](https://doi.org/10.1021/ed4008765).
- Treagust D. F., (1988), Development and use of diagnostic tests to evaluate teacher candidates' misconceptions in science, *Int. J. Sci. Educ.*, **10**(2), 159–169, DOI: [10.1080/0950069880100204](https://doi.org/10.1080/0950069880100204).
- Vosniadou S., (2013), *International handbook of research on conceptual change*, 2nd edn, Routledge, DOI: [10.4324/9780203154472](https://doi.org/10.4324/9780203154472).
- Wu H. K., (2003), Linking the microscopic view of chemistry to real-life experiences: intertextuality in a high-school science classroom, *Sci. Educ.*, **87**(6), 868–891, DOI: [10.1002/sce.10090](https://doi.org/10.1002/sce.10090).
- Yin R. K., (2014), *Case study research: design and methods*, 5th edn, Sage Publications.

