Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

July 2026. Vol. 14, No. 3
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 2000-2015

Changes in Students' Conception on the Concept of Reaction Order in the Implementation of Conceptual Change Model Supported By Worksheets

***[1]Rosalina Eka Permatasari, [2]Wiwin Putriawati**

[1] Undergraduate Program of Chemistry Education, Universitas Negeri Surabaya, Surabaya, Indonesia

[2] Mathematics Education Department, STKIP Paracendekia NW Sumbawa, Sumbawa Besar, Indonesia

*Corresponding Author e-mail: rosalinapermatasari@unesa.ac.id

Received: May 2026; Revised: June 2026; Published: July 2026

Abstract

Misconceptions about reaction order can persist among preservice chemistry teachers, yet evidence remains limited on how individual conceptions change across successive stages of a conceptual-change intervention. This study explored within-person trajectories during a conceptual change model (CCM) implemented through a conceptual change module (CCMo) and a separate ontology–epistemology–axiology (OEA) worksheet. This exploratory embedded multiple-case study used a qualitative-dominant mixed-methods approach. Six preservice chemistry teachers were purposively selected based on misconception-classified responses to a 16-item four-tier diagnostic test. Qualitative evidence from verification, cognitive-conflict, equilibration, worksheet, and final-reconstruction activities was integrated with descriptive diagnostic-test counts and participant questionnaires. At the first verification, one participant was classified as having a misconception, one as not knowing the concept, and four as knowing the concept; at the second verification, five were classified as having a misconception and one as not knowing the concept. During standard CCMo-supported equilibration, five participants achieved full performance, whereas only one achieved full performance on the OEA worksheet. Final reconstruction was rated sufficient for two participants and poor for four. On the final diagnostic test, five participants showed fewer misconception-classified responses than at baseline, while one participant increased from three to five. These heterogeneous patterns indicate that correct task performance during equilibration did not consistently persist into later reconstruction or diagnostic assessment. The findings provide process-level evidence for individualized remediation and support future controlled studies, but they do not establish the causal effectiveness of the CCM or the added value of the worksheet.

Keywords: Misconception; Reaction order; Conceptual change model; Preservice chemistry teachers; Within-person trajectories

How to Cite: Permatasari, R. E., & Putriawati, W. (2026). The Changes in Students' Conception on the Concept of Reaction Order in the Implementation of Conceptual Change Model Supported By Worksheets. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *14*(3), 2000–2015. https://doi.org/10.33394/j-ps.v14i3.21875

https://doi.org/10.33394/j-ps.v14i3.21875

Copyright© 2026, Permatasari & Putriawati
This is an open-access article under the CC-BY License

INTRODUCTION

Reaction order is a foundational construct in chemical kinetics because it links concentration–time data, mathematical rate laws, and mechanistic explanations of chemical change (Bain & Towns, 2016; Rushton et al., 2014). Scientifically, the exponent of a concentration term must be established from kinetic evidence and cannot ordinarily be inferred from a stoichiometric coefficient in the overall equation, except for an elementary reaction (Marzabal et al., 2018; Rushton et al., 2014). Understanding this distinction requires learners to coordinate macroscopic observations, symbolic equations, graphical trends, particulate models, and mathematical relationships rather than apply a memorized algorithm (Bain & Towns, 2016; Rushton et al., 2014). This coordination is difficult: a four-tier kinetics study found that nearly 70% of the identified alternative conceptions emerged from graph-based items, while interview research showed that students often conflated rate constants with equilibrium constants (Bain et al., 2019; Yan & Subramaniam, 2018).

Common misconceptions include assuming that reaction order equals a stoichiometric coefficient, that rate is directly proportional to every reactant concentration, and that a rate law can be read directly from a balanced equation (Kolomuc & Tekin, 2011; Marzabal et al., 2018). These are not isolated factual errors but fragmented relations among evidence, mathematical representation, and molecular explanation that may remain stable across contexts (Tümay, 2016; Vosniadou, 2019). The problem is especially consequential for preservice chemistry teachers because their content conceptions shape later explanations, assessments, and responses to students' reasoning (Wu & Yezierski, 2023; Yakmaci-Guzel, 2013). Therefore, reconstructing reaction-order understanding is a fundamental teacher-education problem rather than a narrow exercise in correcting calculations (Bain & Towns, 2016; Marzabal et al., 2018).

Evidence indicates that misconceptions in chemical kinetics persist from secondary education into undergraduate study and can also be retained by prospective or practicing chemistry teachers (Cakmakci, 2010; Kolomuc & Tekin, 2011; Yan & Subramaniam, 2018). Conventional instruction may leave these ideas intact when it emphasizes equation manipulation without confronting links among experimental data, rate laws, and mechanisms (Marzabal et al., 2018; Rushton et al., 2014). The diagnostic state of the art has progressed from ordinary multiple-choice tests to multi-tier instruments that combine answers, reasons, and confidence judgments to distinguish misconceptions from incomplete knowledge or guessing (Diani et al., 2019; Yan & Subramaniam, 2018). The instructional state of the art includes model-based derivation of rate laws, inquiry and analogy, conceptual-change and refutation texts, simulations, and cognitive-conflict tasks that explicitly elicit and challenge prior conceptions (Addido et al., 2022; Kaya & Geban, 2012; Özmen & Naseriazar, 2018; Rushton et al., 2014; Sinatra & Broughton, 2011). Prior intervention studies, however, measured different outcomes and seldom connected them as individual trajectories.

Kaya and Geban (2012) compared two classes using a pretest–posttest Reaction Rate Concept Test and reported group-level achievement differences, without stage-by-stage individual evidence or a delayed assessment. Addido et al. (2022) combined one-group pre–post scores, open-ended responses, and Likert ratings; although phase-related responses were examined, the conclusions remained primarily aggregated and did not determine whether correct task performance persisted into later reconstruction. Özmen and Naseriazar (2018) administered pretest, posttest, and delayed concept tests, thereby addressing retention, but analyzed retention mainly through experimental–control group scores rather than within-person movement across intervention stages. Shallcross et al. (2024) examined participation and course performance associated with smart worksheets, not the mechanism or stability of conceptual reconstruction.

These studies also did not incorporate an explicitly separated ontology–epistemology–axiology phase (Addido et al., 2022; Kaya & Geban, 2012; Özmen & Naseriazar, 2018; Shallcross et al., 2024). In Posner's framework, conceptual change requires dissatisfaction with an existing conception and acceptance of a new conception as intelligible, plausible, and fruitful (Posner et al., 1982). However, cognitive conflict is not sufficient because learners may ignore, reinterpret, temporarily accept, or only partially integrate discrepant information (Adams et al., 2024; Limón, 2001; Olaogun & Hunsu, 2025). Contemporary research consequently treats conceptual change as context-sensitive and dependent on prior knowledge, epistemic cognition, motivation, confidence, and instructional support, implying that group averages can conceal divergent individual trajectories (DeGlopper & Stowe, 2024; Ikenyirimba & Talanquer, 2026; Olaogun & Hunsu, 2025).

Taken together, prior work provides evidence about intervention-level outcomes but leaves limited evidence of how an individual preservice chemistry teacher's reaction-order conception changes across verification, cognitive conflict, equilibration, and final reconstruction (Bain & Towns, 2016; Olaogun & Hunsu, 2025). In the present study, the conceptual change model (CCM) is implemented through a conceptual change module (CCMo), followed by a distinct ontology–epistemology–axiology (OEA) worksheet phase. An earlier study by the same research group used a three-scientific-questions-enhanced CCM to describe stage-by-stage changes in four participants' conceptions of acid–base indicators (Permatasari et al., 2024). The present study therefore does not claim that individual case tracing or philosophy-of-science prompts are wholly new. Instead, it extends that earlier work to the conceptually different and mathematically demanding topic of reaction order, examines six cases, makes the CCMo and OEA worksheet phases analytically distinct, and explicitly compares apparently correct performance during equilibration with final reconstruction and diagnostic-assessment evidence.

This comparison is important because the same anomalous evidence can produce rejection, peripheral change, temporary task compliance, or genuine restructuring (Adams et al., 2024; Kang et al., 2004; Olaogun & Hunsu, 2025). Individualized remediation is therefore warranted when learners differ in prior knowledge, confidence, cognitive conflict, and capacity to coordinate representations (Kang et al., 2004; Vosniadou, 2019). The OEA worksheet has a defined conceptual rationale: ontological questions clarify what kind of empirical construct reaction order is, epistemological questions emphasize how rate laws are justified from evidence, and axiological questions connect the concept with explanatory and practical value (DeGlopper & Stowe, 2024; Talanquer, 2026; Tümay, 2016). Nevertheless, the noncomparative design cannot establish that the OEA worksheet adds benefit beyond the CCMo, that OEA instruction is more effective than another approach, or that the intervention is causally effective (Creswell, 2009; Fraenkel & Wallen, 2012; Shallcross et al., 2024). The defensible novelty is therefore the documentation of heterogeneous, within-person conceptual trajectories across several stages of a conceptual-change intervention, including evidence of where understanding consolidates, fragments, or remains resistant (Ikenyirimba & Talanquer, 2026; Posner et al., 1982; Tümay, 2016).

Accordingly, this research is framed as an exploratory embedded multiple-case study of six purposively selected preservice chemistry teachers who displayed misconception-classified responses concerning reaction order (Creswell, 2009; Fraenkel & Wallen, 2012). The individual participant is the unit of analysis, followed by cross-case comparison of initial conceptions, conflict responses, CCMo-supported equilibration, OEA worksheet performance, final reconstruction, and diagnostic-assessment patterns (Creswell, 2009). Participants' evaluations of the CCMo and OEA worksheet are treated only as evidence of usability and acceptability, not as evidence of conceptual learning or intervention effectiveness (Creswell, 2009; Fraenkel & Wallen, 2012). The scope is restricted to one concept, six selected participants, and a sequential intervention without a control or worksheet-free comparison condition; therefore, the findings support descriptive and analytical transferability rather than population-level generalization or causal effectiveness claims (Creswell, 2009; Fraenkel & Wallen, 2012). Within these limits, the study contributes a transparent map of heterogeneous conceptual-change trajectories that may inform individualized remediation and future comparative testing of the worksheet's added value (DeGlopper & Stowe, 2024; Ikenyirimba & Talanquer, 2026; Wu & Yezierski, 2023). The study addresses the following research questions:

1. What initial reaction-order conceptions were demonstrated by the six participants?
2. How did each participant respond across misconception verification, cognitive-conflict, CCMo-supported equilibration, OEA worksheet, and final-reconstruction stages?
3. To what extent was correct performance during equilibration retained in the final reconstruction and diagnostic assessment?

4. How did participants evaluate the usability and acceptability of the CCMo and OEA worksheet?

METHOD

Research design

This study used a qualitative-dominant embedded multiple-case design supported by descriptive quantitative diagnostic data (Creswell, 2009; Fraenkel & Wallen, 2012). Each participant constituted one case, and the primary analytic focus was the sequence of written and observed responses across the conceptual change model (CCM) stages. Item-level counts from the four-tier diagnostic test and structured questionnaire ratings were embedded as secondary descriptive evidence. Within-case analysis traced each participant's trajectory, and cross-case analysis compared similarities, divergences, and negative cases across the six trajectories.

The design was nonexperimental. Participants were purposively selected rather than randomly assigned, no control or comparison condition was used, and every participant completed the conceptual change module (CCMo) before the ontology–epistemology–axiology (OEA) worksheet. Consequently, the quantitative results were used only to support description and comparison; they were not treated as evidence of causal effectiveness or of the worksheet's added effect. The previous O1–O6–X time-series diagram was removed because the observations represented non-equivalent intervention stages rather than equivalent repeated pretests and posttests.

Participants

The accessible group consisted of 23 prospective chemistry teachers enrolled in the Chemistry Education Undergraduate Program, Faculty of Mathematics and Natural Sciences, Surabaya State University, in the 2023 intake and in their fifth semester when data were collected in April 2026. Participants had taken a chemical kinetics course in their fourth semester prior to screening.

Eligibility required students to (a) complete the 16-item screening test, (b) demonstrate at least the minimum number or percentage of misconception-classified responses, (c) be available for all intervention and assessment stages, and (d) provide written informed consent. Exclusion criteria were incomplete screening, absence from one or more stages, withdrawal, or other criteria. From the eligible students, six cases (P1–P6) were selected because of an exact purposive-selection rationale, such as maximum variation in misconception burden or complete participation. Neutral P-codes replaced the original working codes to avoid confusion between a participant code and the abbreviation MK for misconception. Their baseline misconception-classified item counts ranged from 3 to 11; this range describes the selected cases but does not replace the prespecified eligibility threshold.

Participants were 20-21 years old; gender distribution was female, if ethically appropriate to report. Recruitment was conducted through Participation was voluntary and course grades. Written informed consent was obtained for timing and procedures. Data collection took place in the second week of April. Ethics approval was granted by the committee, institution, approval number, and date. The lecturer was also/not also a member of the research team; if roles overlapped, the following safeguards were used to reduce coercion and analytical bias. Participant codes were used in all notes and reports.

Instruments and measurement

The primary intervention materials were the CCMo and the separate OEA worksheet. The CCMo elicited prior conceptions, presented two cognitive-conflict tasks, supported equilibration through explanatory material and nine questions, and prompted final reconstruction. The OEA worksheet contained ontology, epistemology, axiology, and self-internalization prompts. Because all participants completed the materials in the same sequence, worksheet performance was analyzed as one stage in the trajectory and not as evidence that the worksheet improved learning.

The Reaction Sequence Diagnostic Instrument contains 16 four-level items representing four elements of a concept: label, definition, essential characteristics, and examples/non-examples (Herron et al., 1977). Each item measures answer accuracy, answer confidence, reasoning accuracy, and reasoning confidence. Table 1 presents all 16 possible combinations. Classification is performed at the item level. A participant's misconception load is the sum of the 16 items classified as misconceptions (MCs), rather than a binary label based on a single item. Participants were selected if they demonstrated misconceptions on at least 30% of the concept items tested.

Table 1. Four-tier item-classification rules

Answer accuracy	Answer confidence	Reason accuracy	Reason confidence	Classification
Correct	Confident	Correct	Confident	Knows the concept (TK)
Correct	Confident	Correct	Not confident	Knows the concept (TK)
Correct	Not confident	Correct	Confident	Knows the concept (TK)
Correct	Not confident	Correct	Not confident	Knows the concept (TK)
Incorrect	Not confident	Incorrect	Not confident	Does not know the concept (TTK)
Incorrect	Not confident	Correct	Not confident	Does not know the concept (TTK)
Incorrect	Not confident	Correct	Confident	Does not know the concept (TTK)
Correct	Not confident	Incorrect	Not confident	Does not know the concept (TTK)
Correct	Confident	Incorrect	Not confident	Does not know the concept (TTK)
Correct	Not confident	Incorrect	Confident	Does not know the concept (TTK)
Correct	Confident	Incorrect	Confident	Does not know the concept (TTK)
Incorrect	Confident	Incorrect	Confident	Misconception (MK)
Incorrect	Confident	Incorrect	Not confident	Misconception (MK)
Incorrect	Confident	Correct	Not confident	Misconception (MK)
Incorrect	Confident	Correct	Confident	Misconception (MK)
Incorrect	Not confident	Incorrect	Confident	Misconception (MK)

Five experts evaluated the Conceptual Change Model (CCMo), the OEA worksheet, the four-tier diagnostic test, and the cognitive conflict response questionnaire. The expert panel consisted of two professors in chemistry education, one professor in chemistry, and two associate professors specializing in chemistry education and educational assessment, each with more than ten years of experience in chemistry education research, instructional design, and assessment development. The experts evaluated the instruments in terms of relevance, clarity, representativeness, alignment with the research objectives, language appropriateness, and content accuracy. A five-point Likert rating scale was used (1 = very poor, 2 = poor, 3 = fair, 4 = good, and 5 = excellent). Content validity was established through expert judgment, and the ratings were summarized using the mode. All items obtained a modal score of 5, indicating that the experts consistently judged every item as highly appropriate for measuring the intended constructs. Accordingly, all items met the predetermined acceptance criterion of a modal score of 5 and were retained. Only minor editorial revisions were made based on the experts' comments, including improving wording clarity, refining scientific terminology, and enhancing the consistency of instructions across the instruments. Expert judgment was used solely as evidence of content validity and was not interpreted as evidence of instrument reliability.

The reliability of the expert validation was established using Borich's (2003) Percentage of Agreement (PA). The Percentage of Agreement represents the level of agreement between expert raters by comparing the scores assigned by one validator with those assigned by another (Borich, 2003, p. 285). The PA was calculated using the following formula:

$$\text{Percentage Agreement (PA)} = \left(1 - \frac{A-B}{A+B}\right) x\ 100\%$$

Where A (the higher score assigned by the validators) and B (the lower score assigned by the validators).

The analysis revealed a content validity agreement of 88.9%, a construct validity agreement of 88.0%, and a language validity agreement of 95.0%. The reliability analysis demonstrated high inter-rater agreement, with Percentage of Agreement values of 90.0% for content validity, 91.6% for construct validity, and 92.0% for language validity, indicating that the instrument achieved a satisfactory level of reliability.

Overall, the findings indicate that the developed instrument is appropriate for use, demonstrating very high levels of content, construct, and language validity, as well as strong reliability. All Percentage of Agreement values exceeded the minimum criterion of 75% recommended by Borich (2003), indicating substantial agreement among the expert validators. Therefore, the use of the mode (Mo) as the basis for determining the final validity decision was considered appropriate and well supported by the high level of inter-rater agreement.

Cognitive conflict levels were determined using an adapted version of the Cognitive Conflict Questionnaire developed by Lee et al. (2003). The questionnaire consisted of 14 items administered during two conflict creation stages of the conceptual change model. Each item was scored using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The cognitive-conflict score was calculated using the mean score formula:

$$CC = \frac{\sum X}{N}$$

where CC is the cognitive-conflict score, ΣX is the total score of all items, and N is the number of items. The cognitive-conflict levels were interpreted as very low (1.00–1.80), low (1.81–2.60), medium (2.61–3.40), high (3.41–4.20), and very high (4.21–5.00). An overall cognitive-conflict level was assigned to each participant based on the mean score, while item-level responses were retained for further analysis to describe specific cognitive-conflict patterns across conceptual-change activities.

The final reconstruction level was determined based on students' responses in the final reconstruction phase of the Conceptual Change Model (CCMo). In this phase, students were required to answer 12 evaluation items related to the target concepts. The final reconstruction score was calculated as the proportion of correctly answered items relative to the total number of items. For example, the student achieved a score of 67% by correctly answering 8 out of 12 items (8/12 × 100). This result indicates that the student had achieved partial conceptual reconstruction; however, the reconstruction was not considered complete because a full reconstruction level requires correct responses to all evaluation items (100%).

Students' evaluations of the CCM were concluded based on questionnaire data and written responses provided by the students. The questionnaire results indicated that students perceived the CCMo and worksheet as highly supportive learning tools. Furthermore, students' written responses in the CCMo and worksheet did not indicate any significant difficulties in following the stages of the conceptual change model. The students' opinions were predominantly characterized by agreement, with a score of 4 (agree) across the 16 indicators evaluated.

Most students stated that the learning materials presented in the CCMo and worksheet were organized into small and specific learning units aligned with the stages of conceptual change. This structure facilitated students in reconstructing their incorrect conceptions and supported their engagement throughout the conceptual change strategy aimed at reducing misconceptions. However, several aspects of the CCMo still require further improvement, such as providing more examples or contextual problems related to everyday life (an axiological perspective) and revising the use of lengthy narrative explanations, which require students to read and interpret the texts more carefully.

The positive written responses were consistent with students' levels of agreement in the questionnaire. This consistency indicates that students who used the CCMo and worksheet provided objective evaluations regarding the relevance and consistency of the developed learning materials. Students' perceptions were also consistent with previous research findings. Diakidoy and Kendeou (2001) reported that individuals demonstrated significantly better conceptual understanding when learning through conceptual change texts compared with standard textbooks. Similarly, Jensen and Finley (as cited in Limon, 2001) reported the effectiveness of conceptual change approaches in improving conceptual understanding. Furthermore, the findings are supported by Kosterelioglu and Dagdelen (2015), who demonstrated that conceptual change strategies can effectively reduce students' misconceptions.

Procedure

Data collection was conducted during the second week of April 2026, consisting of three consecutive sessions. In the first session, students completed a baseline diagnostic test to identify their initial conceptions and misconceptions related to the target chemistry concepts. In the second session, students participated in learning activities using the Conceptual Change Model (CCM)-based module (CCMo) and worksheet. In the third session, students completed the final diagnostic test to evaluate changes in their conceptual understanding after the intervention.

Table 2. Intervention and data-collection sequence

Stage	**Activity and evidence collected**	**Duration and interval**	**Administration controls**
1. Screening	Administration of the 16-item four-tier diagnostic test; item-level TK, TTK, and MK classifications recorded.	Conducted in the first session of the second week of April 2026; approximately 60 minutes.	Conducted in the first session of the second week of April 2026; approximately 60 minutes. The interval before Stage 2 was one week.
2. First verification	Participant judged the statement that reaction order can be determined from stoichiometric coefficients and reported confidence.	Approximately 2-5 minutes during the initial CCMo implementation session.	Approximately 2-5 minutes during the initial CCMo implementation session.
3. Second verification	Participant explained or defined reaction order in their own words; conceptual status was coded.	Approximately 2-5 minutes.	Approximately 5 minutes.
4. First cognitive-conflict task	Participant interpreted scenario-based experimental data and completed the cognitive-conflict measure.	Approximately 5 minutes.	Approximately 5 minutes.
5. Second cognitive-conflict task	Participant compared the prior response with a scientific explanation and completed the second conflict rating.	Approximately 5-10 minutes.	The researcher monitored the activity and provided only procedural clarification. Participants were encouraged to reflect on inconsistencies between their initial conceptions and scientific explanations.
6. Standard CCMo equilibration	Participant read the CCMo explanation and answered nine reaction-order questions.	Approximately 5-10 minutes.	Only procedural and reading assistance was permitted. The researcher clarified instructions or terminology related to task

Stage	Activity and evidence collected	Duration and interval	Administration controls
			completion but did not provide conceptual coaching or direct answers.
7. OEA worksheet activity	Participant completed ontology, epistemology, axiology, and self-internalization prompts.	20 minutes	Participants worked individually. They could review the CCMo explanation provided in the worksheet but could not access their previous diagnostic responses. Assistance was limited to procedural clarification.
8. Final reconstruction	Participant generated a definition, completed an understanding check, and applied the concept; responses were scored with the reconstruction rubric.	8-10 minutes	Responses were completed independently. The researcher did not provide conceptual assistance during reconstruction.
9. Final diagnostic test	The 16-item four-tier instrument was administered and final TK, TTK, and MK item counts were calculated.	60 minutes	The final diagnostic test used parallel items with equivalent conceptual targets but different wording and item arrangement to minimize recall effects. Students were not allowed to access their previous answers or learning records during testing.
10. Response questionnaire	Closed-ended usability/acceptability ratings and written comments about the CCMo and OEA worksheet were collected.	Approximately 15–20 minutes after the final diagnostic test.	Approximately 15–20 minutes after the final diagnostic test.

Data analysis, integration, and trustworthiness

The unit of coding was each participant's response to a diagnostic-test item or a CCMo/OEA intervention prompt at a specific learning stage. Coding was conducted primarily deductively using predefined categories derived from the conceptual framework, including knows the concept (TK), does not know the concept (TTK), misconception (MK), cognitive-conflict level, CCMo and OEA performance, and final conceptual-reconstruction level. Each participant constituted a case and served as the primary unit of analysis. Responses were first examined chronologically within each case and were subsequently compared across cases to identify patterns of conceptual continuity, change, fragmentation, temporary correct performance, and regression. Emerging reasoning patterns that were not adequately represented by the predefined categories were identified inductively and incorporated into the analysis.

In addition, emerging reasoning patterns from students' written explanations and verbal responses were analyzed inductively and incorporated into the coding framework to capture variations in students' conceptual change processes. The coding framework included category definitions, decision rules, and examples of representative responses. The initial coding scheme was developed based on the conceptual change theory and revised after an initial review of participant responses to ensure clarity and consistency.

Two researchers independently coded the qualitative data. The first researcher performed the initial coding, and the second researcher independently reviewed the coding results. The coding agreement was calculated using percentage agreement, which was determined by dividing the number of agreements by the total number of coding decisions and multiplying by 100.

Any differences between coders were resolved through discussion and consensus by referring to the coding criteria and original participant responses. After agreement was achieved, the finalized coding framework was applied to all remaining data. For within-case analysis, evidence for each participant was ordered chronologically from screening through final assessment in a case matrix. The analysis identified points of conceptual continuity, change, fragmentation, temporary correct performance, and regression. For cross-case analysis, the six matrices were compared stage by stage to identify shared and divergent patterns. Contradictory and negative cases were retained explicitly, including the participant whose misconception-classified item count increased and cases in which full CCMo equilibration performance was not retained at final reconstruction.

Quantitative analysis was conducted descriptively. For each participant and assessment occasion, the numbers and percentages of items classified as target knowledge (TK), teaching knowledge/confidence (TTK), and misconception knowledge (MK) were calculated. The percentage of each category was obtained by dividing the number of items within each classification by the total number of diagnostic items and multiplying by 100. Questionnaire responses regarding the usability and acceptability of the CCMo module and worksheet were summarized using item frequencies, percentages, and mean scores for each indicator. Students' open-ended comments were analyzed qualitatively and summarized into content categories representing perceived strengths, difficulties, and suggestions for improving the CCMo and worksheet. The response questionnaire consisted of 16 indicators rated on a five-point Likert scale. The frequency distribution of each response category was calculated, and the mean score for each indicator was used to identify students' overall perceptions of the CCMo and worksheet.

Qualitative and quantitative evidence was integrated through joint displays and narrative comparison. Diagnostic counts were compared with verification responses, conflict ratings, CCMo and OEA task performance, and final reconstruction for the same participant. Questionnaire data were integrated only with the usability and acceptability question and were not used as evidence of conceptual learning. Triangulation was conducted by comparing four sources of evidence within each case: (1) diagnostic classifications (TK, TTK, and MK profiles), (2) written responses produced during CCMo and OEA worksheet activities, (3) observation and documentation records of students' engagement and conceptual-change processes, and (4) questionnaire responses regarding the usability and acceptability of the CCMo module and worksheet.

Table 3. Alignment of research questions, evidence, and analysis

Research question	Data source and indicator	Analytical technique	Output
RQ1. Initial conceptions	Screening test and two verification tasks; item-level TK/TTK/MK classifications and written definitions.	Descriptive counts plus within-case coding of initial reasoning.	Six initial conceptual profiles.
RQ2. Stage-by-stage responses	Verification, two conflict tasks, CCMo equilibration, OEA worksheet, and final reconstruction.	Chronological within-case matrix followed by stage-based cross-case comparison.	Individual trajectories and cross-case patterns.

Research question	Data source and indicator	Analytical technique	Output
RQ3. Retention of correct performance	CCMo equilibration, OEA performance, final reconstruction, and final diagnostic test.	Within-person concordance/discordance comparison; explicit analysis of negative and contradictory cases.	Evidence of consolidation, fragmentation, temporary performance, or regression.
RQ4. Usability and acceptability	Closed-ended questionnaire ratings and open-ended comments.	Descriptive analysis using item frequencies, percentages, and mean scores, complemented by qualitative content categorization of students' comments.	Usability and acceptability profile, kept separate from learning claims.

RESULTS AND DISCUSSION

Initial diagnostic profiles

The baseline four-tier diagnostic test showed substantial variation across the six cases. Misconception-classified responses ranged from 3 of 16 items for P4 to 11 of 16 items for P6. Because the diagnostic categories were assigned at the item level, each participant could simultaneously show knows-the-concept (TK), does-not-know-the-concept (TTK), and misconception (MK) responses across different items.

Verification and stage-by-stage trajectories

The first verification did not classify all participants as having a misconception. P3 was classified as MK, P1 as TTK, and P2, P4, P5, and P6 as TK. At the second verification, P2–P6 were classified as MK and P1 as TTK. These results show that the two verification tasks elicited different forms of knowledge: judging a presented statement and generating or explaining a conception in the participant's own words.

The intervention-stage evidence was organized as a participant-by-stage trajectory matrix rather than a radar chart. This representation preserves the categorical nature of the evidence and avoids treating unlike stages as equivalent numerical scales.

Comparable case summaries

P1 began with 7 MK-classified responses. Both verification tasks were classified as TTK. Conflict ratings varied from low to very high in the first task and were high in the second. P1 answered all nine standard CCMo equilibration questions correctly, but OEA worksheet performance was incomplete. Final reconstruction was sufficient rather than full (67%), and MK-classified responses decreased only from 7 to 6. This case indicates that correct performance during supported equilibration was not fully retained in later generative or diagnostic tasks.

P2 began with 8 MK-classified responses. The first verification was TK, whereas the second was MK. Both conflict ratings were very low. Standard CCMo equilibration was full, OEA worksheet performance was incomplete, and final reconstruction was poor. MK-classified diagnostic responses decreased from 8 to 4. The pattern combines descriptive diagnostic improvement with weak final reconstruction, suggesting unstable or task-dependent understanding.

P3 began with 9 MK-classified responses, pending confirmation against the original item-level record. Both verification tasks were MK. Conflict was low in the first task and medium in the second. Standard CCMo equilibration was full, OEA worksheet performance was incomplete, and final reconstruction was poor. MK-classified responses decreased from 9 to 4. The available evidence therefore shows partial improvement without complete reconstruction.

P4 began with the lowest misconception burden, 3 MK-classified responses. The first verification was TK and the second was MK. Conflict was high in both tasks, yet standard CCMo equilibration and OEA worksheet performance were incomplete. Final reconstruction was sufficient, but MK-classified responses increased from 3 to 5. P4 is the principal negative case and shows that high reported conflict was neither sufficient for full task performance nor predictive of diagnostic improvement.

P5 began with 9 MK-classified responses. The first verification was TK and the second was MK, while both conflict ratings were very low. P5 was the only participant with full performance in both standard CCMo equilibration and the OEA worksheet. Nevertheless, final reconstruction was poor, although MK-classified responses decreased from 9 to 4. This discordance suggests that successful completion of structured tasks did not guarantee a stable, self-generated explanation.

P6 began with the highest misconception burden, 11 MK-classified responses. The first verification was TK and the second was MK. Conflict was medium in the first task and very low in the second. Standard CCMo equilibration was full, OEA worksheet performance was incomplete, and final reconstruction was poor. MK-classified responses decreased from 11 to 8, indicating partial but limited descriptive change.

Final diagnostic changes

Table 4. Changes in four-tier diagnostic classifications

Participant	Pre TK	Pre TTK	Pre MK	Post TK	Post TTK	Post MK	Change in MK	Pattern
P1	3	6	7	5	5	6	−1	Decrease
P2	4	4	8	7	5	4	−4	Decrease
P3	2	5	9	7	5	4	−5	Decrease
P4	8	5	3	6	5	5	+2	Increase
P5	3	4	9	8	4	4	−5	Decrease
P6	1	4	11	4	4	8	−3	Decrease

Five participants showed fewer MK-classified responses at the final diagnostic assessment, whereas P4 increased from 3 to 5. No participant eliminated misconception-classified responses. Across all cases, the total number of MK-classified responses decreased descriptively from 47 to 31, but this aggregate change is not evidence of causal effectiveness because the study had no comparison condition and did not control for repeated testing, maturation, or other alternative explanations.

Usability and acceptability

Questionnaire and written-comment data indicated that participants generally regarded the CCMo and OEA worksheet as usable and acceptable. Responses were dominated by agreement ratings, and participants reported that the materials were organized into manageable units that followed the intervention sequence. They also recommended more everyday-life examples for the axiology component and shorter narrative passages. These perceptions describe usability and acceptability only; they are not evidence that the materials produced conceptual change.

Discussion

Heterogeneous trajectories and unstable transfer

The central finding is not uniform improvement but heterogeneity within and across participants. Five cases showed full performance during standard CCMo equilibration, yet none achieved full final reconstruction, and all retained misconception-classified responses. This divergence is consistent with the view that conceptual change can be partial, context-sensitive, and unstable rather than an all-or-none replacement of an earlier conception (Tümay, 2016; Vosniadou, 2019; Olaogun & Hunsu, 2025). The baseline distributions in Figure 1 demonstrate that the participants entered the intervention with markedly different conceptual profiles. P6 displayed the heaviest misconception burden (11 of 16 items), whereas P4 showed the lightest burden (3 of 16 items). The coexistence of TK, TTK, and MK responses within the same participant also indicates that initial understanding was fragmented rather than uniformly correct or incorrect. Baseline misconception load therefore characterizes the complexity of each case, but it should not be treated as a deterministic predictor of subsequent reconstruction.

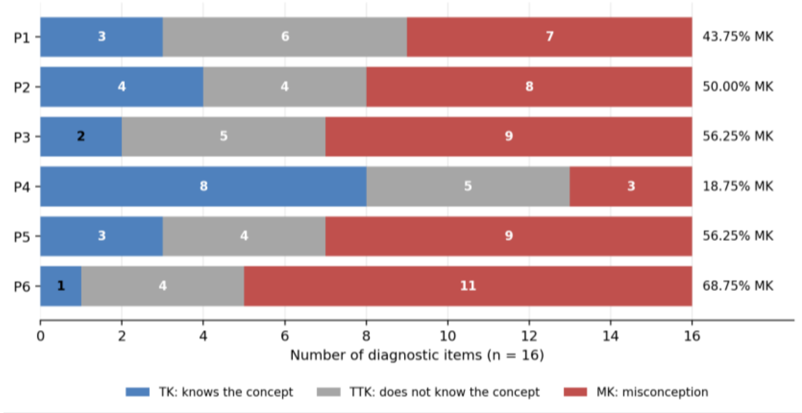
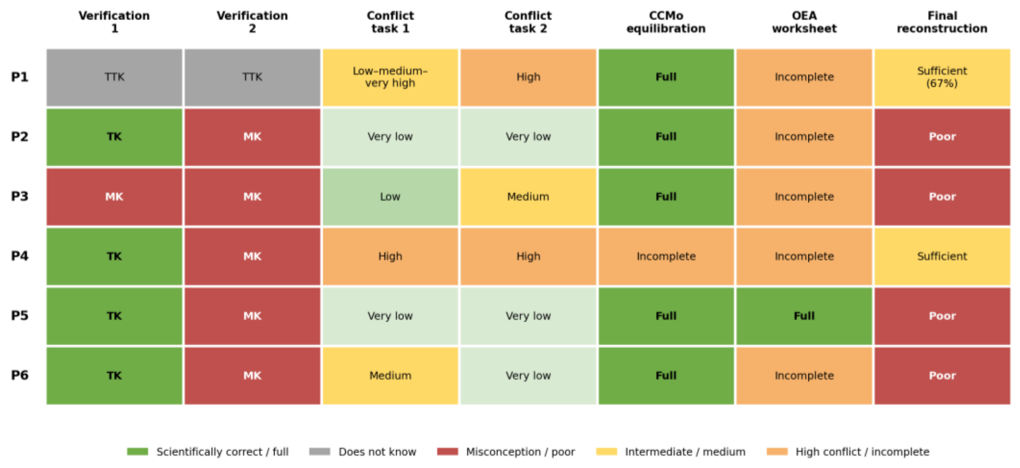


Figure 1. Baseline four-tier diagnostic profiles across the six participants

Several nonexclusive explanations may account for the decline from supported task performance to final reconstruction. Participants may have temporarily complied with the task structure, corrected surface features without reorganizing the underlying conception, remembered wording from the instructional text, or succeeded on recognition-based prompts but struggled to generate an explanation independently. Other possibilities include difficulty coordinating symbolic, graphical, and experimental representations; unstable metaconceptual awareness; weak transfer from OEA prompts to diagnostic items; and insufficient time for consolidation. The present evidence cannot distinguish among these mechanisms, but the within-person discordances identify them as testable hypotheses for future studies.

Figure 2 further shows that the trajectories did not follow a simple progression from cognitive conflict to stable conceptual reconstruction. Five participants reached full CCMo equilibration, but only two achieved sufficient final reconstruction and none achieved full reconstruction. P5 provides the clearest example of task-dependent performance: full completion of both the CCMo and OEA stages was followed by poor final reconstruction. Conversely, P4 showed incomplete performance in both structured stages but reached a sufficient reconstruction rating. These cross-stage reversals indicate that success on scaffolded tasks and success in independently generating a scientific explanation are related but non-equivalent forms of evidence.



Note. TK = knows the concept; TTK = does not know the concept; MK = misconception; CCMo = conceptual change module; OEA = ontology–epistemology–axiology.

Figure 2. Participant-by-stage conceptual trajectory diagram

Cognitive conflict was not a monotonic predictor

The cases do not support a simple proposition that stronger cognitive conflict produces more successful accommodation. P2, P5, and P6 showed very low conflict at one or both stages but achieved full standard equilibration. P1 showed high conflict and full standard equilibration but only sufficient final reconstruction. P4 showed high conflict in both tasks, incomplete equilibration, and an increase in MK-classified responses. These counterexamples align with cautions that cognitive conflict may be ignored, reinterpreted, or resolved without deep restructuring (Limón, 2001; Kang et al., 2004). Conflict should therefore be treated as one condition within a broader process involving prior knowledge, epistemic cognition, motivation, confidence, and instructional support (DeGlopper & Stowe, 2024; Olaogun & Hunsu, 2025).

Cognitive conflict and cognitive load are distinct constructs. Cognitive conflict concerns perceived inconsistency between an existing conception and new evidence, whereas cognitive load concerns demands on working-memory resources. This study measured conflict ratings but did not measure cognitive load. Accordingly, cognitive-load terminology has been removed from the interpretation of the conflict data.

The OEA worksheet cannot be interpreted as an enrichment effect

Only P5 achieved full OEA worksheet performance, compared with five participants who achieved full standard CCMo equilibration. This pattern does not demonstrate that the worksheet enriched understanding. The worksheet may have assessed more demanding generative dimensions, especially the axiology component; its items may have been less aligned with the core reaction-order construct; the language may have increased reading demands; or its scoring criteria may have been more stringent. It is also possible that the worksheet provided no added conceptual benefit. Because every participant completed the CCMo before the worksheet and no worksheet-free condition was included, the study cannot separate these possibilities or estimate an added effect.

Poor worksheet performance should therefore not be attributed automatically to participant failure. Future instrument studies should examine the construct alignment, linguistic demand, scoring severity, and dimensionality of the ontology, epistemology, and axiology prompts before the worksheet is evaluated as an instructional component.

Diagnostic change and causal limits

The final diagnostic counts provide descriptive evidence of change: five participants had fewer MK-classified responses, one had more, and none eliminated them. These results are compatible with the resistance and possible re-emergence of misconceptions reported in the conceptual-change literature (Barke et al., 2009; Duit, 2009; Ibrahim, 2012). However, the study cannot attribute the observed changes to the CCM, CCMo, or OEA worksheet. The appropriate formulations are that changes occurred within the intervention sequence, were associated with participants' completion of the stages, and varied by case—not that the intervention caused equilibration or was proven effective.

Limitations

The findings are limited by six purposively selected cases from one institution and one chemistry concept. The absence of random assignment, a control or comparison condition, and a worksheet-free condition prevents causal inference and isolation of the worksheet's contribution. Reuse of identical or closely related diagnostic items may have produced testing or memory effects. The design included no delayed posttest, so stability beyond the immediate assessment is unknown. Possible researcher–instructor role overlap may have influenced participation or responses, and questionnaire findings are subject to self-report bias. Instrument reliability, coding agreement, and the scoring thresholds for conflict and reconstruction must be fully reported from the original study records. Finally, the language and reading demands of the CCMo and OEA worksheet may have affected performance independently of reaction-order understanding.

CONCLUSION

This multiple-case study documented heterogeneous conceptual-change trajectories across six preservice chemistry teachers. Misconception-classified diagnostic responses decreased for five participants but increased for one, and every participant retained some misconception-classified responses at the final assessment. Five participants demonstrated full performance during standard CCMo equilibration, but this success did not consistently transfer to final reconstruction: two cases were rated sufficient and four poor. OEA worksheet performance was incomplete for five of the six participants. High cognitive conflict did not consistently correspond with stronger reconstruction or diagnostic improvement, and several low-conflict cases performed well during standard equilibration. These patterns indicate that conceptual change was nonlinear, task-dependent, and sometimes temporary. Participants' positive usability and acceptability ratings should not be interpreted as evidence of conceptual effectiveness. The study therefore contributes process-level descriptions of within-person change and negative cases, while controlled and delayed-assessment research is needed to test intervention effects and the worksheet's added value.

RECOMMENDATION

Future research should test these process-level findings with larger and more diverse samples, a comparison condition, a worksheet-free CCMo condition, and a delayed diagnostic assessment. Studies should also evaluate the reliability and construct alignment of the four-tier diagnostic test, cognitive-conflict measure, reconstruction rubric, and OEA worksheet; distinguish recognition from generative explanation; and examine whether language and reading demands affect performance. Such designs would permit stronger conclusions about retention, transfer, and any added contribution of the OEA worksheet.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the students who participated in this research. They also thank the Undergraduate Program of Chemistry Education of the Faculty of Mathematics and Natural Sciences, Unesa, and all parties who supported the data collection and preparation of this research.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rosalina Eka Permatasari	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Wiwin Putriawati		✓				✓		✓	✓	✓	✓	✓		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

Written informed consent was obtained from all participants before their participation in the study. Participants were informed about the purpose, procedures, voluntary nature of participation, confidentiality measures, and their right to withdraw from the study without penalty.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Adams, A., & Barth-Cohen, L. (2024). Exploring student sensemaking when engaging with anomalous data. *CBE—Life Sciences Education, 23*(4), ar63. https://doi.org/10.1187/cbe.24-08-0208

Addido, J., Burrows, A., & Slater, T. (2022). The effect of the conceptual change model on conceptual understanding of electrostatics. *Education Sciences, 12*(10), 696. https://doi.org/10.3390/educsci12100696

Bain, K., Rodriguez, J.-M. G., & Towns, M. H. (2019). Investigating student understanding of rate constants: When is a constant "constant"? *Journal of Chemical Education, 96*(8), 1571–1577. https://doi.org/10.1021/acs.jchemed.9b00005

Bain, K., & Towns, M. H. (2016). A review of research on the teaching and learning of chemical kinetics. *Chemistry Education Research and Practice, 17*(2), 246–262. https://doi.org/10.1039/C5RP00176E

Barke, H.-D., Hazari, A., & Yitbarek, S. (2009). *Misconceptions in chemistry: Addressing perceptions in chemical education*. Springer. https://doi.org/10.1007/978-3-540-70989-3

Borich, G. D. (2003). *Observation skills for effective teaching* (4th ed.). Merrill/Prentice Hall.

Cakmakci, G. (2010). Identifying alternative conceptions of chemical kinetics among secondary school and undergraduate students in Turkey. *Journal of Chemical Education, 87*(4), 449–455. https://doi.org/10.1021/ed8001336

Creswell, J. W. (2009). *Research design: Pendekatan penelitian kualitatif, kuantitatif, dan mixed* (A. Fawaid, Trans.). Pustaka Pelajar.

DeGlopper, K. S., & Stowe, R. L. (2024). Modeling students' epistemic cognition in undergraduate chemistry courses: A review. *Chemistry Education Research and Practice, 25*(3), 594–612. https://doi.org/10.1039/D3RP00348E

Diakidoy, I. A. N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. *Learning and Instruction, 11*(1), 1–20. https://doi.org/10.1016/S0959-4752(00)00011-6

Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-tier diagnostic test with certainty of response index on the concepts of fluid. *Journal of Physics: Conference Series, 1155*(1), 012078. https://doi.org/10.1088/1742-6596/1155/1/012078

Duit, R. (2009). *Bibliography: Students' and teachers' conceptions and science education*. Leibniz Institute for Science and Mathematics Education, University of Kiel. https://archiv.leibniz-ipn.de/stcse/

Fraenkel, J. R., & Wallen, N. E. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Herron, J. D., Cantu, L. L., Ward, R., & Srinivasan, V. (1977). Problems associated with concept analysis. *Science Education, 61*(2), 185–199. https://doi.org/10.1002/sce.3730610210

Ibrahim, M. (2012). *Konsep, miskonsepsi, dan cara pembelajarannya*. Unesa University Press.

Ikenyirimba, O. J., & Talanquer, V. (2026). Conceptual reframing in action: How students build and revise predictions about a simple chemical reaction. *Chemistry Education Research and Practice, 27*(3), 953–967. https://doi.org/10.1039/D5RP00440C

Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education, 34*(1), 71–96. https://doi.org/10.1023/B:RISE.0000021001.77568.b3

Kaya, E., & Geban, Ö. (2012). Facilitating conceptual change in rate of reaction concepts using conceptual change oriented instruction. *Education and Science, 37*(163), 216–225. https://doi.org/10.15390/ES.2012.1014

Kolomuç, A., & Tekin, S. (2011). Chemistry teachers' misconceptions concerning concept of chemical reaction rate. *International Journal of Physics and Chemistry Education, 3*(2), 84–101. https://doi.org/10.51724/ijpce.v3i2.194

Kösterelioğlu, İ., & Dağdelen, O. (2015). Effect of conceptual change texts for overcoming misconceptions in the "People and Management" unit. *International Electronic Journal of Elementary Education, 8*(1), 99–112.

Lee, G., Kwon, J., Park, S. S., Kim, J. W., Kwon, H. G., & Park, H. K. (2003). Development of an instrument for measuring cognitive conflict in secondary-level science classes. *Journal of Research in Science Teaching, 40*(6), 585–603. https://doi.org/10.1002/tea.10099

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change:

A critical appraisal. *Learning and Instruction, 11*(4–5), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Marzabal, A., Delgado, V., Moreira, P., Barrientos, L., & Moreno, J. (2018). Pedagogical content knowledge of chemical kinetics: Experiment selection criteria to address students' intuitive conceptions. *Journal of Chemical Education, 95*(8), 1245–1249. https://doi.org/10.1021/acs.jchemed.8b00296

Olaogun, O. P., & Hunsu, N. J. (2025). A systematic review of factors that predict and mediate conceptual change. *European Journal of Psychology of Education, 40*(2), Article 58. https://doi.org/10.1007/s10212-025-00959-1

Özmen, H., & Naseriazar, A. (2018). Effect of simulations enhanced with conceptual change texts on university students' understanding of chemical equilibrium. *Journal of the Serbian Chemical Society, 83*(1), 121–137. https://doi.org/10.2298/JSC161222065O

Permatasari, R. E., Suyono, & Yuanita, L. (2024). Investigating students' conception of acid-base indicators using three scientific questions-enhanced conceptual change model. *Jurnal Pendidikan MIPA, 25*(1), 341–350. https://doi.org/10.23960/jpmipa/v25i1.pp341-350

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Rushton, G. T., Criswell, B. A., McAllister, N. D., Polizzi, S. J., Moore, L. A., & Pierre, M. S. (2014). Charting an alternate pathway to reaction orders and rate laws in introductory chemistry courses. *Journal of Chemical Education, 91*(1), 66–73. https://doi.org/10.1021/ed3006743

Shallcross, D. E., Davies-Coleman, M. T., Lloyd, C., Heslop, B., Eastman, J., Baldwin, T., Thistlethwaite, I., & Franklin, S. D. (2024). Smart worksheets and their positive impact on a second-year chemical kinetics course. *Journal of Chemical Education, 101*(4), 1735–1740. https://doi.org/10.1021/acs.jchemed.3c00934

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly, 46*(4), 374–393. https://doi.org/10.1002/RRQ.005

Talanquer, V. (2026). The importance of fostering conceptual reframing in chemistry education. *Chemistry Education Research and Practice, 27*(2), 693–700. https://doi.org/10.1039/D5RP00421G

Tümay, H. (2016). Reconsidering learning difficulties and misconceptions in chemistry: Emergence in chemistry and its implications for chemical education. *Chemistry Education Research and Practice, 17*(2), 229–245. https://doi.org/10.1039/C6RP00008H

Vosniadou, S. (2019). The development of students' understanding of science. *Frontiers in Education, 4*, Article 32. https://doi.org/10.3389/feduc.2019.00032

Wu, M.-Y. M., & Yezierski, E. J. (2023). Secondary chemistry teacher learning: Precursors for and mechanisms of pedagogical conceptual change. *Chemistry Education Research and Practice, 24*(1), 245–262. https://doi.org/10.1039/D2RP00160H

Yakmaci-Guzel, B. (2013). Preservice chemistry teachers in action: An evaluation of attempts for changing high school students' chemistry misconceptions into more scientific conceptions. *Chemistry Education Research and Practice, 14*(1), 95–104. https://doi.org/10.1039/C2RP20109G

Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chemistry Education Research and Practice, 19*(1), 213–226. https://doi.org/10.1039/C7RP00143F