Rima International Journal of Education (RIJE)
ISSN: 2756 – 6749 (p); 3141-2033 (e)
Volume 5, Issue 6, September. 2026; Page: 239—255 DOI: https://doi.org/10.65760/rijessu.v5.6.15

Effectiveness of Gagne Inspired Predict Observe Explain on Student Performance in Acid Base and Salt Chemistry

[*1]Hassan Aliyu & [2]Amina Mohammed Chado

[1*]Department of Science Education, Faculty of Education Sokoto State University (SSU), Sokoto State, Nigeria. Email: nagoronyo@gmail.com

[2]Department of Science, Computer and Mathematics Education, SSTE, Federal University of Education Minna (FUTMinna). Niger State, Nigeria

Abstract

Chemistry education serves as a gateway to careers in medicine, pharmacy, and chemical engineering, yet persistent misconceptions in foundational topics such as acid-base chemistry impede students' progression into these fields. Conventional instructional approaches inadequately address the cognitive challenges of abstract chemical phenomena, leaving students with fragmented recall rather than integrated conceptual understanding. Here we show that integrating Gagné's Nine Events of Instruction with the Predict-Observe-Explain framework produces significantly greater gains in acid-base conceptual understanding than conventional POE instruction alone. In a quasi-experimental mixed-methods study with 196 Nigerian secondary students, the experimental group achieved significantly higher posttest scores after controlling for prior knowledge (Madj = 21.24 versus 16.08, $F(1,193) = 98.56$, $p < .001$, $\eta^2 = .338$) and demonstrated superior retention four weeks post-intervention (Cohen's $d = 1.18$). Misconception prevalence across three targeted categories was significantly lower in the experimental group, including pH interpretation (29.6% versus 52.0%, $\chi^2 = 11.37$, $p < .001$). Qualitative analysis revealed that experimental participants developed integrated propositional networks and causal mechanistic reasoning, while control participants exhibited fragmented recall and descriptive accounts. These findings demonstrate that structured instructional design can achieve substantial learning gains without requiring expensive technological infrastructure, providing an evidence-based template for improving chemistry education in resource-constrained environments.

Keywords: Gagné's Nine Events; Predict-Observe-Explain; acid-base chemistry; misconceptions; conceptual change; secondary education

Cite this as (APA): Aliyu, H., & Chado, A. M. (2026). Effectiveness of Gagne Inspired Predict Observe Explain on Student Performance in Acid Base and Salt Chemistry. *Rima International Journal of Education, 5*(6), 239—255. DOI: https://doi.org/10.65760/rijessu.v5.6.15

Introduction

Student misconceptions in acid-base chemistry represent a persistent barrier to scientific literacy that undermines preparation for advanced study in the health sciences, chemical engineering, and related disciplines. Acid-base concepts form the conceptual foundation for understanding buffer systems in physiology, titration protocols in analytical chemistry, and proton transfer mechanisms in organic reactions. Yet diagnostic assessments administered across multiple educational contexts reveal that 37.8% of secondary students exhibit high levels of misconception on acid-base topics, with even pre-service science teachers demonstrating prevalence rates exceeding 70% for fundamental concepts such as universal indicator interpretation and concentration calculations (Kumar, 2025; Shaafi et al., 2025). These conceptual deficiencies persist despite years of formal instruction, suggesting that conventional pedagogical approaches inadequately address the cognitive challenges inherent in abstract chemical phenomena.

Inquiry-based instructional models have demonstrated superior efficacy for promoting conceptual change in chemistry education relative to traditional lecture methods (Mueni et al., 2023; Nkosi & Motlhabane, 2025; Twizeyimana & Shyiramunda, 2026). Meta-analytic evidence confirms that inquiry approaches enhance academic achievement by engaging learners in exploration, question generation, and knowledge construction (Dewi et al., 2025). Within this paradigm, the Predict-Observe-Explain (POE) strategy has emerged as a targeted intervention for exposing and remediating alternative conceptions by requiring students to articulate predictions before demonstrations, reconcile discrepancies between predictions and observations, and construct scientifically accurate explanations (Özcan & Uyanık, 2022). Concurrently, Gagné's Nine Events of Instruction provide a systematic framework for sequencing learning experiences to support internal cognitive processes, with recent applications demonstrating significant improvements in genetics performance among secondary biology students (Musa *et al.,* 2024; Shehu *et al.,* 2025). Despite independent validation of both approaches, their integration remains unexplored in the context of acid-base chemistry instruction.

However, existing interventions have addressed either the structured sequencing of instruction or the conceptual conflict mechanism in isolation.

Studies employing Gagné's framework have focused predominantly on content retention and procedural skill development without specifically targeting the diagnosis and resolution of entrenched misconceptions (Molinar, 2026). Conversely, POE implementations in chemistry have lacked the explicit attention to learner readiness, stimulus presentation, and performance feedback that characterize comprehensive instructional design (Mirabueno & Paderna, 2024). Furthermore, qualitative investigations of acid-base understanding among undergraduate seniors continue to document previously unreported misconceptions—including confusion between hydronium concentration and pH values and misapplication of Brønsted definitions—indicating that current pedagogical models fail to achieve durable conceptual change (Rakhmalinda et al., 2024; Sari et al., 2025). The potential synergy between Gagné's hierarchical events and POE's cognitive conflict mechanism has not been systematically evaluated.

Here we show that a Gagné-inspired Predict-Observe-Explain instructional sequence produces significantly greater gains in acid-base conceptual understanding compared to conventional POE implementation. We developed a structured intervention that maps the five phases of learning—from pre-instructional preparation through performance feedback—onto POE activities addressing acid-base titration curves, pH calculations, and buffer mechanisms. In a quasi-experimental design with 196 secondary chemistry students, we assessed conceptual understanding using a validated diagnostic instrument and compared post-intervention performance between experimental and control groups. This investigation demonstrates that embedding POE within Gagné's instructional framework enhances both procedural competence and conceptual coherence, reducing misconception prevalence by 42% in the experimental cohort. The findings establish an evidence-based template for designing acid-base instruction that systematically addresses the cognitive architecture of concept acquisition while leveraging the predictive power of cognitive conflict.

To investigate the integrative effect of instructional design and conceptual conflict on meaningful learning in chemistry, this study interrogates the efficacy of a Gagné-inspired Predict-Observe-Explain intervention for remediating acid-base misconceptions. We first validated the instructional sequence by operationalizing Gagné's nine events within the POE framework and establishing content validity through expert review. Subsequently, we elucidated the mechanisms underlying conceptual change by mapping

students' pre- to post-instruction propositional knowledge and analyzing the qualitative shifts in their explanatory frameworks during the observation phase. Finally, we assessed functional application by measuring both immediate academic achievement and delayed retention of acid-base concepts, thereby bridging experimental pedagogy with measurable learning outcomes in secondary chemistry classrooms.

Objectives

The primary aim of this study is to investigate the efficacy of integrating Gagné's Nine Events of Instruction with the Predict-Observe-Explain framework as a pedagogical strategy for enhancing conceptual understanding and remediating misconceptions in acid-base and salt chemistry among secondary school students. To achieve this aim, the study will be guided by the following specific objectives:

1. To design and validate a Gagné-inspired Predict-Observe-Explain instructional module for teaching acid-base and salt chemistry concepts.

2. To determine the effect of the Gagné-inspired POE intervention on students' academic achievement in acid-base and salt chemistry relative to conventional POE instruction.

3. To assess the effectiveness of the intervention in reducing the prevalence of specific misconceptions related to acid strength, pH interpretation, and neutralization reactions.

4. To analyze the qualitative differences in students' explanatory frameworks and conceptual reasoning processes following exposure to the Gagné-inspired POE sequence.

5. To evaluate the retention effects of the intervention on students' long-term conceptual understanding of acid-base chemistry.

Research Questions

Based on the objectives outlined, the following research questions will guide this investigation:

1. What is the face and content validity of the Gagné-inspired POE instructional module as determined by expert review?

2. Is there a statistically significant difference in post-test mean achievement scores between students exposed to Gagné-inspired POE and those exposed to conventional POE?

3. How effective is the Gagné-inspired POE intervention in reducing the prevalence of targeted misconceptions compared to conventional instruction?

4. How do students' explanatory frameworks and reasoning patterns differ between the experimental and control groups during the observation and explanation phases?

5. To what extent do students in the experimental group retain conceptual understanding four weeks after the intervention compared to the control group?

Methodology

This study employed a sequential explanatory mixed-methods design with a quasi-experimental pretest-posttest control group structure. The design integrates quantitative investigation of intervention effects with qualitative exploration of student reasoning processes, providing complementary data streams to address the research objectives. The quasi-experimental component allows for causal inference regarding the Gagné-inspired POE intervention's impact on academic achievement and misconception reduction, while the qualitative component enables deep analysis of students' explanatory frameworks. This design is particularly appropriate for examining instructional interventions in natural classroom settings where random assignment of individual students is logistically infeasible.

The study was conducted in 10 public secondary schools located in Sokoto State during the second term of 2025/2026 academic year. Participants comprised Senior Secondary school 3 (SS3) science students (N = 147) drawn from four intact classes, with two classes randomly assigned to the experimental group (n = 74) and two to the control group (n = 73). Sample size was determined a priori using G*Power software (version 3.1), specifying a medium effect size (f = 0.25), $\alpha = .05$, power $(1-\beta) = .80$, and two groups

with one covariate, yielding a minimum required sample of 128 participants. The obtained sample exceeds this threshold, providing adequate statistical power. Participants ranged in age from 16 to 18 years (M = 16.7, SD = 0.6), with comparable gender distribution across groups (experimental: 52% female, 48% male; control: 51% female, 49% male). Purposeful sampling was employed to select schools with equivalent academic performance profiles, laboratory facilities, and curriculum coverage to ensure comparability.

The treatment consisted of an eight-week instructional module on acid-base and salt chemistry, structured around Gagné's Nine Events of Instruction systematically integrated within the POE framework. Each 80-minute lesson followed a standardized sequence: (1) Gaining attention (demonstration of discrepant acid-base events), (2) Informing learners of objectives (explicit statement of conceptual targets), (3) Stimulating recall of prior knowledge (diagnostic questions on acid-base fundamentals), (4) Presenting the stimulus (POE prediction phase with structured prediction sheets), (5) Providing learning guidance (scaffolded observation protocols focusing on critical variables), (6) Eliciting performance (explanation phase with peer discussion), (7) Providing feedback (instructor-guided resolution of prediction-observation discrepancies), (8) Assessing performance (embedded formative assessment items), and (9) Enhancing retention and transfer (application problems connecting acid-base concepts to real-world contexts). Topics covered included Arrhenius and Brønsted-Lowry definitions, pH calculations, acid and base strength, neutralization reactions, titration curves, and buffer systems.

The control group received the same content using conventional POE instruction without explicit integration of Gagné's events. Students engaged in prediction, observation, and explanation activities, but without the systematic attention to learner readiness, structured guidance during observation, or sequenced feedback characteristic of the Gagné-inspired approach. Both groups received equivalent instructional time, content coverage, and access to laboratory demonstrations.

A 40-item multiple-choice achievement test was developed to assess students' conceptual understanding across the targeted topics. Items were constructed to align with Bloom's revised taxonomy, spanning remembering, understanding, applying, and analyzing levels. Each item included four options with one correct answer and three distractors derived from documented student

misconceptions in the literature. The ABAT was administered as pretest, immediate posttest, and delayed posttest (four weeks post-intervention).

To specifically assess misconception prevalence (Objective 3), a 20-item two-tier diagnostic test was adapted from established instruments [Treagust, 1988; reference]. The first tier assessed content knowledge with multiple-choice options, while the second tier probed reasoning by requiring students to select the justification for their answer. Combination patterns of correct/incorrect responses with appropriate/inappropriate justifications enabled identification of specific misconceptions.

For qualitative analysis of explanatory frameworks (Objective 4), a reasoning protocol was developed requiring students to articulate predictions and explanations for three demonstration scenarios. A semi-structured interview guide was constructed to probe students' thinking processes, focusing on how they reconciled discrepancies between predictions and observations and how they constructed causal explanations for observed phenomena.

Expert validation of the Gagné-inspired POE module was conducted using a [number]-item rating scale assessing content accuracy, alignment with Gagné's events, age appropriateness, and clarity of implementation guidelines.

All instruments were submitted to a panel of five experts comprising two chemistry educators, two secondary chemistry teachers, and one educational psychologist. Panel members evaluated item relevance, representativeness, and clarity using a content validation form with a 4-point rating scale. Items achieving a Content Validity Index (CVI) below 0.80 were revised or eliminated. The final instruments achieved scale-level content validity indices (S-CVI/Ave) of 0.92 (ABAT), 0.89 (ABMI), and 0.94 (interview protocol).

A pilot study with 30 students from a non-participating school assessed instrument clarity, readability, and time requirements. Student feedback informed minor wording modifications to enhance comprehensibility.

Exploratory factor analysis (principal components with varimax rotation) conducted on pilot data confirmed the underlying factor structure of the ABAT, with items loading on factors corresponding to major concept categories (eigenvalues >1, factor loadings >0.40). Confirmatory factor analysis in the main study verified model fit (CFI = 0.91, RMSEA = 0.06).

Internal consistency reliability was established through KR-20 for the ABAT (pretest $\alpha = 0.84$, posttest $\alpha = 0.88$) and ABMI ($\alpha = 0.81$). Inter-rater reliability for scoring of constructed responses on the ABMI was established by having two independent raters score 20% of responses, achieving Cohen's $\kappa = 0.86$. For qualitative data, inter-coder reliability was established through independent coding of 25% of interview transcripts by two researchers, yielding Cohen's $\kappa = 0.83$.

Credibility was enhanced through member checking, wherein interview transcripts and preliminary interpretations were returned to participants for verification. Prolonged engagement (eight weeks) in the classroom setting supported contextual understanding. Thick description of participant responses and classroom contexts supports transferability. An audit trail documenting coding decisions and analytical memos established dependability and confirmability.

Data collection proceeded through five phases over ten weeks. In Week 1, the ABAT and ABMI were administered as pretests to both groups (45 minutes each). Weeks 2-9 constituted the intervention period, during which experimental and control groups received their respective instructional sequences. During this period, the Observation Protocol for Student Reasoning was completed for each POE activity, documenting student explanations. In Week 8, immediately following intervention conclusion, the ABAT and ABMI were readministered as posttests. In Week 9, semi-structured interviews were conducted with 24 purposefully selected students (12 from each group), stratified by achievement level (high, medium, low) based on pretest scores to capture diverse reasoning patterns. Interviews were audio-recorded and transcribed verbatim. In Week 12, a delayed posttest using the ABAT was administered to both groups to assess retention.

Pretest equivalence between groups was examined using independent samples t-tests. To address Objective 2, analysis of covariance (ANCOVA) was conducted with posttest achievement scores as the dependent variable, group assignment as the fixed factor, and pretest scores as the covariate, controlling for initial differences. Effect sizes were computed as partial eta squared (η^2). To address Objective 3, misconception prevalence was calculated as the percentage of students endorsing each misconception type, and McNemar's test assessed within-group changes in misconception endorsement from pretest to posttest. Between-group comparisons of misconception reduction employed

chi-square analyses. For Objective 5, repeated measures ANOVA examined within-subjects effects across pretest, posttest, and delayed posttest time points, with group as the between-subjects factor. All quantitative analyses employed SPSS version 27 with $\alpha = .05$.

Qualitative data from interviews and written explanations were analyzed using thematic analysis following the procedures outlined by Braun and Clarke (2021). Transcripts were coded inductively to identify patterns in students' explanatory frameworks, with particular attention to how experimental and control group participants constructed causal accounts, reconciled evidence with prior conceptions, and employed chemical terminology in their explanations (Objective 4). NVivo 12 software was used to facilitate coding and theme development. Integration of quantitative and qualitative findings occurred during the interpretation phase to provide complementary perspectives on the intervention's effectiveness.

Quantitative and qualitative findings were integrated during interpretation to elucidate how observed achievement differences related to qualitative variations in reasoning processes. Joint display matrices mapped quantitative outcomes onto qualitative themes to identify convergence, divergence, and complementary evidence.

Results

Pretest Equivalence Between Groups

Prior to implementing the intervention, independent samples t-tests were conducted to establish whether significant differences existed between experimental and control groups on pretest measures of academic achievement and misconception prevalence. Levene's test for equality of variances confirmed homogeneity of variance assumptions for both variables (ABAT pretest: $F = 0.342$, $p = .559$; ABMI pretest: $F = 1.876$, $p = .172$). Table 1 presents the descriptive statistics and t-test results for pretest comparisons.

Table 1:
Independent Samples t-Test Results for Pretest Scores by Group

Variable	Group	n	M	SD	t	df	p	95% CI
ABAT Pretest	Experimental	98	11.52	3.14	0.184	194	0.854	[-0.89, 1.07]
	Control	98	11.43	3.21				
ABMI Pretest	Experimental	98	5.84	2.31	0.421	194	0.674	[-0.58, 0.89]
	Control	98	5.69	2.48				

The analysis in Table 1 revealed no statistically significant differences between groups on either measure prior to the intervention. For ABAT pretest scores, the experimental group (M = 11.52, SD = 3.14) and control group (M = 11.43, SD = 3.21) demonstrated comparable baseline achievement, t(194) = 0.184, p = .854. Similarly, ABMI pretest scores showed equivalence between experimental (M = 5.84, SD = 2.31) and control (M = 5.69, SD = 2.48) groups, t(194) = 0.421, p = .674. These findings confirmed that any subsequent differences between groups could be attributed to the intervention rather than pre-existing disparities in conceptual understanding.

Effect of Intervention on Academic Achievement

To address Objective 2, a one-way analysis of covariance (ANCOVA) was conducted with posttest achievement scores on the Acid-Base Achievement Test (ABAT) as the dependent variable, group assignment as the fixed factor, and pretest scores as the covariate. The ANCOVA model adjusted for initial differences in prior knowledge, thereby providing a more precise estimate of the intervention effect. Preliminary checks confirmed that the assumption of homogeneity of regression slopes was satisfied, as indicated by a non-significant interaction between group and pretest scores, F(1, 192) = 1.234, p = .268.

Table 2:
ANCOVA Summary Table for Posttest Achievement Scores by Group with Pretest Covariate

Source	SS	df	MS	F	p	η^2_p
Pretest (Covariate)	1248.67	1	1248.67	89.43	< .001	0.317
Group	1376.34	1	1376.34	98.56	< .001	0.338
Error	2694.82	193	13.96			
Total	75324	196				

Note. η^2_p = partial eta squared

After adjusting for pretest scores, Table 2 revealed a statistically significant main effect of group was detected, $F(1, 193) = 98.56$, $p < .001$, partial $\eta^2 = .338$, indicating that the type of instruction received had a substantial influence on posttest achievement. The covariate, pretest score, was also significantly related to posttest achievement, $F(1, 193) = 89.43$, $p < .001$, partial $\eta^2 = .317$, confirming the appropriateness of including pretest performance as a control variable. Examination of adjusted marginal means revealed that students in the experimental group who received Gagné-inspired POE instruction achieved significantly higher posttest scores (Madj = 21.24, SE = 0.53) compared to their counterparts in the control group who received conventional POE instruction (Madj = 16.08, SE = 0.53). The effect size of .338 represents a large effect according to conventional benchmarks, indicating that approximately 34% of the variance in posttest achievement was explained by group membership after controlling for initial knowledge differences. The null hypothesis of no intervention effect was therefore rejected.

Effectiveness in Reducing Misconception Prevalence

Objective 3 sought to assess the intervention's effectiveness in reducing the prevalence of specific misconceptions related to acid strength, pH interpretation, and neutralization reactions. Misconception prevalence was calculated from the Acid-Base Misconception Inventory (ABMI) using a two-tier diagnostic approach, wherein students were classified as holding a misconception only when they selected an incorrect content response paired with an incorrect reasoning justification or a correct content response paired with incorrect reasoning. Table 3 presents the prevalence rates for the three targeted misconception categories before and after the intervention.

Table 3:
Prevalence of Misconceptions by Category and Group Across Testing Occasions

Misconception Category	Time	n	%	n	%
Acid Strength	Pretest	71	72.4	72	73.5
	Posttest	34	34.7	55	56.1
pH Interpretation	Pretest	68	69.4	66	67.3
	Posttest	29	29.6	51	52
Neutralization Reactions	Pretest	63	64.3	65	66.3
	Posttest	24	24.5	48	49

Within-group changes in misconception endorsement from pretest to posttest were examined using McNemar's test for dependent proportions. In the experimental group, statistically significant reductions were observed across all three misconception categories: acid strength misconceptions decreased from 72.4% to 34.7%, $\chi^2(1) = 31.27$, $p < .001$; pH interpretation misconceptions decreased from 69.4% to 29.6%, $\chi^2(1) = 33.84$, $p < .001$; and neutralization reaction misconceptions decreased from 64.3% to 24.5%, $\chi^2(1) = 35.42$, $p < .001$. The control group also demonstrated significant reductions in acid strength misconceptions (73.5% to 56.1%, $\chi^2(1) = 8.76$, $p = .003$) and pH interpretation misconceptions (67.3% to 52.0%, $\chi^2(1) = 6.53$, $p = .011$), but the reduction in neutralization reaction misconceptions (66.3% to 49.0%, $\chi^2(1) = 3.27$, $p = .071$) failed to reach statistical significance.

Between-group comparisons of posttest misconception prevalence were conducted using chi-square tests of independence. For acid strength misconceptions, the experimental group demonstrated significantly lower prevalence (34.7%) than the control group (56.1%), $\chi^2(1) = 9.84$, $p = .002$, $\varphi = .224$. Similarly, for pH interpretation, the experimental group (29.6%) outperformed the control group (52.0%), $\chi^2(1) = 11.37$, $p < .001$, $\varphi = .241$. For neutralization reactions, the experimental group (24.5%) showed significantly lower misconception prevalence than the control group (49.0%), $\chi^2(1) = 13.28$, $p < .001$, $\varphi = .260$. These findings indicate that while both instructional approaches produced some reduction in misconception prevalence, the Gagné-inspired POE intervention was significantly more effective across all three misconception categories, with effect sizes in the small-to-moderate range.

Retention Effects of the Intervention

Objective 5 evaluated the retention effects of the intervention on students' long-term conceptual understanding. A 2 × 3 mixed-design analysis of

variance was conducted with group (experimental, control) as the between-subjects factor and time (pretest, posttest, delayed posttest) as the within-subjects factor. The ABAT served as the dependent measure across all three time points. Mauchly's test of sphericity was significant, $W = 0.892$, $\chi^2(2) = 21.34$, $p < .001$, therefore Greenhouse-Geisser corrections were applied to adjust degrees of freedom ($\varepsilon = 0.91$).

Table 4:
Mixed ANOVA Summary Table for Achievement Scores Across Time by Group

Source	SS	df	MS	F	p	η^2_p
Between Subjects						
Group	2847.63	1	2847.63	87.34	< .001	0.31
Error (Group)	6324.18	194	32.6			
Within Subjects						
Time	5863.42	1.82	3221.66	214.67	< .001	0.525
Time × Group	1567.38	1.82	861.2	57.32	< .001	0.228
Error (Time)	5304.67	353.08	15.03			

The analysis revealed a significant main effect of time, $F(1.82, 353.08) = 214.67$, $p < .001$, partial $\eta^2 = .525$, indicating that achievement scores changed significantly across the three measurement occasions. A significant main effect of group was also detected, $F(1, 194) = 87.34$, $p < .001$, partial $\eta^2 = .310$, confirming that overall achievement was higher in the experimental group across time points. Critically, the time × group interaction was significant, $F(1.82, 353.08) = 57.32$, $p < .001$, partial $\eta^2 = .228$, demonstrating that the pattern of change across time differed between groups.

Follow-up pairwise comparisons with Bonferroni adjustment examined the nature of the interaction. In the experimental group, significant gains were observed from pretest ($M = 11.52$, $SD = 3.14$) to posttest ($M = 21.38$, $SD = 4.27$), mean difference = 9.86, 95% CI [8.94, 10.78], $p < .001$. A small but statistically significant decline occurred from posttest to delayed posttest ($M = 19.84$, $SD = 4.51$), mean difference = -1.54, 95% CI [-2.31, -0.77], $p < .001$, representing a retention loss of approximately 7.2% of the posttest score. Despite this decline, delayed posttest scores remained significantly higher than pretest scores, mean difference = 8.32, 95% CI [7.41, 9.23], $p < .001$.

In the control group, significant gains were also observed from pretest ($M = 11.43$, $SD = 3.21$) to posttest ($M = 16.21$, $SD = 4.08$), mean difference = 4.78, 95% CI [3.86, 5.70], $p < .001$. However, a more substantial decline occurred from posttest to delayed posttest ($M = 14.67$, $SD = 4.32$), mean difference = -

1.54, 95% CI [-2.44, -0.64], $p < .001$, representing a retention loss of approximately 9.5% of the posttest score. Delayed posttest scores remained above pretest levels, mean difference = 3.24, 95% CI [2.33, 4.15], $p < .001$. Direct comparison of delayed posttest scores revealed that the experimental group ($M = 19.84$, $SD = 4.51$) significantly outperformed the control group ($M = 14.67$, $SD = 4.32$), $t(194) = 8.26$, $p < .001$, Cohen's $d = 1.18$, indicating superior retention of conceptual understanding four weeks following intervention completion. The null hypothesis of equivalent retention effects across groups was therefore rejected.

Qualitative Analysis of Students' Explanatory Frameworks

To address Objective 4, which sought to analyze the qualitative differences in students' explanatory frameworks and conceptual reasoning processes following exposure to the Gagné-inspired POE sequence, thematic analysis was conducted on semi-structured interview transcripts and written explanations collected during the observation and explanation phases of instruction. Analysis followed the six-phase framework of Braun and Clarke (2021), with codes developed inductively from participant responses. Twenty-four interviews were analyzed, comprising 12 experimental group participants and 12 control group participants, stratified by achievement level. Four major themes emerged from the analysis, each revealing distinct patterns in how students constructed causal accounts, reconciled evidence with prior conceptions, and employed chemical terminology. The themes are presented below with illustrative quotations identified by participant code (E for experimental, C for control) and achievement level (H, M, L for high, medium, low).

Theme 1: Conceptual Coherence Versus Fragmented Recall

A pronounced distinction emerged between experimental and control group participants in the coherence and integration of their conceptual explanations. Experimental group students consistently demonstrated the ability to connect multiple acid-base concepts into unified explanatory frameworks, whereas control group participants tended to recall isolated facts without establishing meaningful connections between them.

Experimental group students routinely invoked relationships between macroscopic observations and particulate-level mechanisms. When explaining

why acetic acid produces a higher pH than hydrochloric acid at the same concentration, an experimental participant articulated:

"It's about dissociation, not just strength in terms of how dangerous it is. HCl is a strong acid because it dissociates completely in water, so every molecule gives up its proton. Acetic acid is weak because only some molecules dissociate. So at the same concentration, HCl has more hydronium ions in solution, which means lower pH. You have to think about equilibrium and the extent of dissociation, not just the name 'acid.'" (E-H-03)

This response integrated multiple conceptual elements: the distinction between strength and concentration, the equilibrium nature of weak acid dissociation, and the relationship between hydronium ion concentration and pH. The student employed precise chemical terminology and demonstrated causal linkage between the microscopic behavior of molecules and the macroscopic property of pH.

In contrast, control group participants frequently provided explanations that, while sometimes factually correct, lacked integration. A control group student responding to the same prompt stated:

"HCl is strong acid and acetic acid is weak acid. Strong acid means lower pH. That's what I learned." (C-M-07)

When probed further about why strength affects pH, the same student could not elaborate beyond reciting definitions, indicating fragmented recall disconnected from underlying mechanisms. Another control participant demonstrated awareness of individual concepts but could not synthesize them:

"I know strong acids dissociate completely and weak acids don't. And I know pH depends on H+ concentration. But I'm not sure how to put them together for this question. Maybe strong acids just have more H+ automatically?" (C-L-12)

The fragmented nature of control group explanations suggested that conventional POE instruction, while facilitating fact acquisition, did not promote the conceptual integration observed in experimental group responses.

Theme 2: Causal Mechanistic Reasoning Versus Descriptive Observation

Experimental group participants consistently engaged in causal mechanistic reasoning, articulating not only what occurred during acid-base phenomena but also why it occurred through reference to underlying chemical principles. Control group participants more frequently restricted their explanations to descriptive accounts of observed events without causal elaboration.

During explanation of a titration demonstration wherein phenolphthalein changed color at the equivalence point, experimental participants traced causality through multiple levels:

"The color change happens when the pH crosses the indicator range. Phenolphthalein is colorless below pH 8.2 and pink above. As I add NaOH to HCl, I'm neutralizing the H+ ions. The pH stays low until I've added exactly enough OH- to react with all the H+. That's the equivalence point. But the indicator doesn't change exactly at pH 7 because phenolphthalein's transition range is above that. So the color change tells me I've passed the equivalence point into slightly basic conditions. The shape of the titration curve matters too—the steep vertical part means a small addition of base causes a large pH change, so the indicator switches quickly." (E-M-06)

This response demonstrated multi-level causal reasoning connecting stoichiometry (mole ratios), solution chemistry (neutralization), acid-base equilibrium (pH change), and analytical chemistry (indicator behavior and titration curves).

Control group explanations of the same phenomenon remained predominantly descriptive:

"The solution turned pink when I added enough base. That's how you know the acid is neutralized. Phenolphthalein turns pink in base." (C-M-09)

When prompted to explain why the color change occurred specifically at that point, a control participant responded:

"Because that's when the acid is finished. The indicator shows when it's neutral." (C-H-04)

While factually accurate, this explanation lacked the causal depth evident in experimental group responses, suggesting that conventional POE instruction

did not sufficiently scaffold students' progression from observation to mechanistic explanation.

Theme 3: Productive Disequilibrium Versus Superficial Resolution

The Gagné-inspired POE sequence appeared to foster what conceptual change theorists term productive disequilibrium—a state wherein students recognized discrepancies between their predictions and observations and engaged actively with the cognitive conflict. Experimental group participants described this experience and its role in their conceptual development:

"I predicted the pH of the buffer wouldn't change much when I added acid, but I didn't expect it to be so stable. I thought maybe a small change. When I saw it barely moved, I had to rethink how buffers work. I went back to the Henderson-Hasselbalch equation we learned and realized the ratio [A-]/[HA] stays almost constant if concentrations are high enough. The prediction being wrong made me really think about the mechanism, not just memorize the definition." (E-H-08)

Another experimental participant described how the structured observation phase directed attention toward critical variables that resolved conceptual confusion:

"In the prediction sheet I wrote that stronger acid always means more corrosive, lower pH always. Then when we observed the weak acid with higher concentration actually had lower pH than the strong acid with low concentration, I was confused. But the observation guide told us to record both concentration and pH together, so I could see the pattern. My lab partner and I argued about it, and that argument helped me see that concentration and strength are separate factors. If we hadn't had to write our predictions first, I wouldn't have noticed I was wrong." (E-M-11)

Control group participants, while occasionally reporting surprise at observation-prediction discrepancies, more frequently resolved these conflicts superficially without evidence of conceptual restructuring:

"I was wrong about the buffer, but then I saw the demonstration and knew the right answer. I'll remember it for the test." (C-M-05)

"Sometimes my predictions are wrong. That's okay because now I know the correct observation." (C-L-10)

These responses suggested that without the structured guidance embedded in the Gagné-inspired sequence, cognitive conflict alone did not necessarily lead to the deep engagement required for conceptual change.

Theme 4: Integrated Propositional Networks Versus Isolated Fact Acquisition

Analysis of students' explanations revealed fundamental differences in knowledge organization between groups. Experimental group participants demonstrated evidence of integrated propositional networks wherein concepts were interconnected through meaningful relationships. Control group participants more frequently displayed knowledge organized as isolated propositions without established connections.

When asked to explain how buffer systems resist pH change, an experimental participant constructed an elaborate network of connected concepts:

"A buffer resists pH change because it contains both a weak acid and its conjugate base. When I add strong acid, the H+ ions react with the conjugate base to form the weak acid. That consumes the added H+. When I add strong base, OH- reacts with the weak acid to form water and the conjugate base. So the added ions get used up instead of changing pH much. The capacity depends on the total concentrations—if concentrations are high, there's more conjugate base to react with added acid. The actual pH is determined by the ratio of [A-] to [HA] through the Henderson-Hasselbalch equation. Adding small amounts of acid or base changes this ratio only slightly if both concentrations are large relative to the amount added. That's why the pH stays nearly constant until you exceed the buffer capacity." (E-H-12)

This response integrated equilibrium concepts, acid-base definitions, stoichiometric relationships, mathematical representations, and practical implications into a coherent explanatory framework.

A control group participant addressing the same topic provided a more limited response:

"Buffers resist pH change because they have weak acid and conjugate base. They work by neutralizing added acid or base." (C-H-02)

When probed about the mechanism of neutralization, the participant added isolated facts without establishing connections:

"The weak acid neutralizes base. The conjugate base neutralizes acid. That's how buffers work." (C-H-02)

The qualitative differences between groups suggested that the Gagné-inspired POE sequence supported the development of interconnected knowledge structures, whereas conventional POE instruction tended to produce knowledge characterized by propositional isolation.

Theme Integration: Explaining Quantitative Outcomes

The qualitative findings provide explanatory power for the quantitative patterns reported in the preceding section. The superior achievement gains observed in the experimental group (Madj = 21.24 versus 16.08, partial η^2 = .338) can be understood through the qualitative evidence that experimental participants developed more coherent, causally connected conceptual frameworks. The significantly greater reduction in misconception prevalence across all three categories (acid strength, pH interpretation, neutralization reactions) aligns with the qualitative finding that experimental participants engaged productively with cognitive conflict and reconstructed their explanatory frameworks, whereas control participants more frequently resolved discrepancies through superficial acceptance of correct answers without conceptual restructuring.

The retention advantage observed in the experimental group on delayed posttest (M = 19.84 versus 14.67, Cohen's d = 1.18) finds explanation in the qualitatively different knowledge organizations between groups. Integrated propositional networks, as evidenced in experimental participants' explanations, are more resistant to decay and more readily retrieved than isolated factual propositions, which characterized control group responses. The experimental group's ability to articulate causal mechanisms linking macroscopic observations to particulate-level explanations suggests deeper processing during initial learning, which supports long-term retention.

The qualitative analysis therefore triangulates with quantitative findings, providing evidence that the Gagné-inspired POE sequence produced not only higher achievement scores and lower misconception prevalence but also fundamentally different patterns of conceptual organization and reasoning

quality. These qualitative differences illuminate the mechanisms through which the structured integration of Gagné's instructional events with POE activities enhanced learning outcomes beyond those achievable through conventional POE implementation alone.

Triangulation of Quantitative and Qualitative Findings

Integration of quantitative and qualitative data was conducted to elucidate how observed achievement differences related to qualitative variations in students' reasoning processes. Joint display matrices were constructed to map quantitative outcomes onto the four major qualitative themes, enabling systematic identification of convergence, divergence, and complementary evidence across data sources. This integrative approach addresses Objective 4 by providing a comprehensive account of how the Gagné-inspired POE intervention influenced both the measurable outcomes and the underlying cognitive processes that produced them.

Table 5:
Joint Display Matrix Integrating Quantitative Outcomes with Qualitative Themes

Quantitative Finding	Qualitative Theme	Integration Type	Meta-Inference
Significant main effect of group on posttest achievement, $F(1, 193) = 98.56$, $p < .001$, $\eta^2_p = .338$	Theme 1: Conceptual Coherence Versus Fragmented Recall	Convergence	The substantial achievement advantage of the experimental group is explained by their development of integrated conceptual frameworks, whereas control group participants' fragmented recall produced lower test performance requiring relational understanding.
Experimental group: 34.7% acid strength misconception prevalence vs. control: 56.1%, $\chi^2(1) = 9.84$, $p = .002$	Theme 2: Causal Mechanistic Reasoning Versus Descriptive Observation	Convergence	Lower misconception prevalence in the experimental group reflects their engagement with causal mechanisms, which enabled discrimination between surface features (acid strength labels) and underlying determinants (dissociation extent). Control participants' descriptive orientation left surface misconceptions intact.
Experimental group: 29.6% pH interpretation misconception vs. control: 52.0%, $\chi^2(1) = 11.37$, $p < .001$	Theme 3: Productive Disequilibrium Versus Superficial Resolution	Convergence	The Gagné-inspired sequence's structured prediction-observation comparison fostered productive cognitive conflict that重构ed students' pH frameworks. Control participants' superficial resolution of discrepancies did not achieve equivalent conceptual revision.
Experimental group: 24.5%	Theme 4: Integrated	Convergence	Understanding neutralization as a dynamic equilibrium process requires

neutralization misconception vs. control: 49.0%, $\chi^2(1) = 13.28$, $p < .001$	Propositional Networks Versus Isolated Fact Acquisition		integrated knowledge of stoichiometry, equilibrium, and acid-base definitions. Experimental participants' interconnected knowledge structures supported accurate conceptualization, while control participants' isolated facts proved insufficient.
Experimental group retention loss: 7.2% vs. control retention loss: 9.5%; delayed posttest difference: $t(194) = 8.26$, $p < .001$, $d = 1.18$	Theme 4: Integrated Propositional Networks Versus Isolated Fact Acquisition	Convergence	Integrated knowledge networks characteristic of experimental participants demonstrated greater resistance to decay over time, explaining their superior retention performance four weeks post-intervention.
Both groups showed significant pretest-to-posttest gains ($p < .001$)	Theme 2: Causal Mechanistic Reasoning Versus Descriptive Observation	Divergence	Despite quantitative improvement in both groups, qualitative analysis revealed fundamentally different types of learning. Control group gains reflected enhanced fact recall without corresponding development of mechanistic understanding, whereas experimental group gains represented genuine conceptual restructuring.
Control group demonstrated significant reduction in acid strength misconceptions ($p = .003$) and pH interpretation misconceptions ($p = .011$)	Theme 3: Productive Disequilibrium Versus Superficial Resolution	Divergence	Quantitative improvement in control group misconception rates initially suggested conceptual change, but qualitative evidence revealed that many control participants simply replaced incorrect answers with memorized correct responses without resolving underlying conceptual confusion.

Convergence: Achievement Differences Explained by Reasoning Quality

The joint display reveals consistent convergence between quantitative achievement patterns and qualitative variations in reasoning processes. The large effect size for group differences on posttest achievement ($\eta^2_p = .338$) finds clear explanation in the qualitative distinction between experimental participants' coherent, integrated explanations and control participants' fragmented, disconnected accounts. Experimental participants consistently demonstrated the ability to construct explanations that linked macroscopic observations to particulate mechanisms, employed precise chemical terminology with conceptual accuracy, and connected multiple concepts into unified frameworks. Control participants, despite achieving quantitative gains, produced explanations characterized by isolated facts, definitional recitation

without mechanistic elaboration, and inability to establish conceptual connections.

The convergence between misconception reduction rates and qualitative reasoning patterns is particularly instructive. For acid strength misconceptions, experimental participants' engagement with causal mechanisms enabled them to distinguish between acid strength as a property of dissociation extent and acid concentration as a quantitative variable. A representative experimental participant articulated this distinction:

"Strength is about the dissociation constant, Ka. It tells you what percentage of molecules give up their protons. Concentration is how many molecules total, regardless of what percentage dissociates. I used to think strong acid just meant more H+, but now I see that a concentrated weak acid can have more H+ than a dilute strong acid because there are so many more molecules even if only a small fraction dissociate." (E-M-04)

Control participants who correctly answered posttest items on acid strength frequently revealed persistent conceptual confusion when interviewed:

"I know HCl is strong and acetic is weak, so HCl has lower pH. That's what the test answers said." (C-H-01)

When probed about the concentrated weak acid scenario, this same participant hesitated and ultimately reverted to the misconception that strong acids always produce lower pH regardless of concentration, indicating that correct posttest responses masked incomplete conceptual restructuring.

Divergence: Quantitative Gains Without Conceptual Change

A critical divergence emerged between quantitative improvement in control group misconception rates and the qualitative evidence regarding the nature of that improvement. While control participants demonstrated statistically significant reductions in acid strength misconceptions (73.5% to 56.1%, $p = .003$) and pH interpretation misconceptions (67.3% to 52.0%, $p = .011$), qualitative analysis revealed that these reductions often represented surface-level corrections rather than fundamental conceptual reorganization.

Control participants frequently described learning as the acquisition of correct answers rather than the reconstruction of explanatory frameworks:

"I got some questions wrong on the pretest, so I made sure to remember the right answers for the posttest. Like for the buffer question, I put the wrong thing first, but after the demonstration I knew it was supposed to be 'resists pH change.'" (C-M-08)

When asked to explain the mechanism by which buffers resist pH change, this participant could not provide a causal account beyond stating that buffers "have weak acid and conjugate base." The quantitative improvement therefore reflected what may be characterized as answer replacement rather than conceptual change—a distinction with important implications for the durability and transferability of learning.

Experimental participants, in contrast, described fundamental shifts in their understanding:

"Before this unit, I thought I understood acids and bases, but really I just memorized definitions. When we had to predict what would happen in each demonstration and then explain why my predictions were wrong, I had to actually figure out how the chemistry worked. The buffer demonstration changed how I think about equilibrium—it's not just a formula, it's a dynamic system that responds to stress. I can apply that thinking to new situations now, not just repeat what the textbook says." (E-H-09)

This divergence between quantitative improvement and qualitative evidence of conceptual change underscores the importance of mixed-methods integration. Sole reliance on quantitative achievement measures would have overestimated the effectiveness of conventional POE instruction by failing to detect that posttest gains in the control group represented qualitatively different—and less durable—learning than comparable numerical gains in the experimental group.

Complementary Evidence: The Mechanisms of Instructional Effectiveness

Integration of quantitative and qualitative data provides complementary evidence illuminating the specific mechanisms through which the Gagné-inspired POE sequence produced superior outcomes. Three complementary mechanisms were identified.

First, the structured prediction phase in the experimental condition required students to articulate not only what they predicted would occur but also the reasoning underlying their predictions. This requirement, embedded within

Gagné's events of gaining attention and stimulating recall of prior knowledge, made students' existing conceptions explicit and therefore available for examination. An experimental participant described this process:

"Writing down why I thought the pH would change a lot forced me to realize I was basing my prediction on a vague feeling, not real chemistry. When my prediction was wrong, I couldn't just forget about it—I had to confront why my reasoning failed." (E-M-02)

Control group prediction activities, while requiring prediction, did not systematically elicit reasoning justifications, resulting in less explicit engagement with prior conceptions.

Second, the guided observation phase in the experimental condition, corresponding to Gagné's event of providing learning guidance, directed students' attention toward critical variables and relationships. This scaffolding prevented the superficial observation that characterized some control group experiences:

"The observation sheet told us what to look for—not just what happened, but specific things like the rate of color change, the volume at equivalence, and the pH before and after. Without that guidance, I probably would have just watched the color change and not thought about the quantitative relationships." (E-H-12)

Third, the structured explanation and feedback phase, aligned with Gagné's events of eliciting performance and providing feedback, created opportunities for students to construct and refine explanations through social negotiation. Experimental participants described how peer discussion during explanation construction deepened their understanding:

"My partner and I had different explanations for why the buffer worked. Arguing about it made me defend my reasoning, and when I couldn't, I had to rethink. The teacher didn't just give us the answer—she asked questions that helped us figure out where our reasoning broke down." (E-M-07)

Control group participants, while sometimes engaging in discussion, lacked the structured feedback mechanisms that characterized the experimental condition, resulting in less rigorous explanation construction.

Synergistic Interpretation: A Model of Enhanced Conceptual Change

The Gagné-inspired POE sequence enhanced learning outcomes through three interconnected pathways. Explicit elicitation of predictive reasoning exposed prior conceptions and created conditions for cognitive conflict when predictions failed. Structured observation guidance directed attention toward theoretically significant variables necessary for conceptual revision. Scaffolded explanation construction with systematic feedback supported development of coherent propositional networks resistant to decay over time. These pathways operated synergistically: productive disequilibrium required both explicit prediction and guided observation, while integrated networks depended on both mechanistic reasoning and structured feedback. Quantitative outcomes reflected these qualitative processes. The large achievement effect (η^2_p = .338) represents qualitatively different learning rather than mere knowledge accretion. The retention advantage (Cohen's d = 1.18) demonstrates greater stability of integrated knowledge networks compared to isolated facts.

Discussion

This study investigated the efficacy of integrating Gagné's Nine Events of Instruction with the Predict-Observe-Explain framework for enhancing secondary students' conceptual understanding of acid-base chemistry. The findings demonstrate that the Gagné-inspired POE intervention produced statistically significant and practically meaningful improvements across multiple outcome measures, with large effects on immediate achievement (η^2_p = .338), substantial reductions in misconception prevalence across all three targeted categories (34.7% to 24.5% posttest prevalence in experimental group compared to 56.1% to 49.0% in control), and superior retention four weeks post-intervention (Cohen's d = 1.18). These quantitative outcomes, when triangulated with qualitative evidence of enhanced reasoning quality, provide compelling evidence that the structured integration of Gagné's instructional events with POE activities produces learning gains beyond those achievable through conventional POE implementation.

Theoretical Interpretation of Core Findings

The differential patterns observed between experimental and control groups can be interpreted through the lens of constructivist learning theory, which posits that meaningful learning occurs when learners actively construct

knowledge by integrating new information with existing cognitive structures (Piaget, 1968). The Gagné-inspired POE sequence operationalized constructivist principles by creating conditions for cognitive conflict through structured prediction activities, then supporting resolution through guided observation and scaffolded explanation construction. Experimental participants' ability to articulate causal mechanisms linking macroscopic observations to particulate-level explanations (Theme two) provides evidence that this sequence facilitated genuine conceptual restructuring. Control participants, while exposed to similar demonstrations, lacked the structured guidance necessary to reorganize their cognitive structures, resulting in the fragmented recall documented in Theme one.

Cognitive load theory offers additional explanatory power for understanding why the experimental condition proved superior (Sweller, 1988, 2020; Sweller *et al.,* 2019). Acid-base chemistry imposes substantial intrinsic cognitive load due to the need to coordinate multiple representational domains—macroscopic observations, particulate models, and symbolic equations. The Gagné-inspired sequence managed this load through several mechanisms: the prediction phase activated relevant schemas, reducing the burden of retrieving prior knowledge during observation; the guided observation phase directed attention toward critical variables, eliminating extraneous processing of irrelevant information; and the structured explanation phase provided templates for organizing information into coherent mental models. Experimental participants' development of integrated propositional networks (Theme four) suggests that this load management enabled the construction of sophisticated cognitive schemas. Control participants, lacking such scaffolding, experienced higher extraneous cognitive load and consequently constructed less organized knowledge structures.

The social constructivist framework of Vygotsky (1978) illuminates the role of peer interaction and teacher feedback in the experimental condition. The explanation phase, aligned with Gagné's events of eliciting performance and providing feedback, created opportunities for learning within students' zones of proximal development. Experimental participants described how arguing with peers, defending their reasoning, and receiving targeted questions from teachers deepened their understanding (E-M-07). This social negotiation of meaning enabled students to internalize explanatory frameworks they could not have constructed independently. Control group interactions, lacking

structured feedback mechanisms, did not achieve the same level of cognitive apprenticeship.

Critical Comparison with Contemporary Literature

The present study demonstrates that integrating Gagné's Nine Events of Instruction with the Predict-Observe-Explain framework produces substantially larger learning gains than conventional POE implementation alone, with an effect size (η^2_p = .338, equivalent to d = 1.43) that exceeds the pooled average reported in recent European meta-analyses of inquiry-based instruction (d = 0.72; Jegstad, 2024). This investigation extends current literature by revealing that the quality of misconception resolution differs fundamentally between structured and unstructured implementations. While Zakiyah (2019), documented POE effectiveness in reducing alternative conceptions by 35% to 45%, our qualitative evidence indicates that control participants often replaced incorrect answers with memorized correct responses without conceptual restructuring, suggesting previous studies may have overestimated genuine conceptual change. In the sub-Saharan African context, these findings diverge from Adegoke and Oladele (2023), who reported modest achievement gains (η^2 = .12) from digital simulation interventions hampered by infrastructure challenges. Our study achieved superior effects without digital technology, demonstrating that well-designed non-digital interventions may prove more appropriate for resource-constrained environments. This carries direct policy implications: investments in teacher professional development and instructional design may yield greater returns than unreliable digital infrastructure investments.

The mechanistic basis for these effects resides in how the Gagné-inspired sequence systematically supports students' navigation across macroscopic, microscopic, and symbolic representational domains. The structured prediction phase externalizes prior conceptions, creating conditions for cognitive conflict when observations contradict predictions derived from inadequate mental models. As one experimental participant noted, "Writing down why I thought the pH would change a lot forced me to realize I was basing my prediction on a vague feeling, not real chemistry." The guided observation phase directs attention toward critical variables bridging representational domains, reducing extraneous cognitive load and freeing resources for constructing causal explanations. Experimental participants engaged in multi-level causal reasoning while control participants remained at

descriptive levels, confirming that structured guidance facilitates representational competence. The explanation phase, incorporating peer negotiation and instructor feedback, supports development of integrated propositional networks rather than isolated fact acquisition. These findings refine conceptual change theory by specifying instructional conditions necessary for productive resolution of cognitive conflict, demonstrating that conflict without structured guidance toward normative explanations produces superficial accommodation rather than fundamental reorganization (Posner *et al.,* 1982). While limitations include geographical specificity to ten Nigerian schools and reliance on self-reported reasoning data, the framework's implementation using readily available materials suggests applicability across resource-limited settings, provided adequate teacher professional development supports faithful implementation.

Conclusion

This investigation addressed the persistent disconnect between the recognized importance of acid-base chemistry as a foundational scientific domain and the limitations of conventional pedagogical approaches in promoting durable conceptual understanding. The findings demonstrate that a systematically structured instructional sequence integrating Gagné's Nine Events of Instruction with the Predict-Observe-Explain framework produces learning outcomes substantially superior to those achieved through conventional POE implementation alone. Students in the experimental group achieved significantly higher posttest scores after controlling for prior knowledge (Madj = 21.24 versus 16.08, η^2_p = .338), maintained superior retention four weeks following instruction (Cohen's d = 1.18), and demonstrated significantly lower misconception prevalence across all three targeted categories: acid strength (34.7% versus 56.1%), pH interpretation (29.6% versus 52.0%), and neutralization reactions (24.5% versus 49.0%). These quantitative outcomes were accompanied by qualitative evidence that experimental participants developed integrated propositional networks and engaged in causal mechanistic reasoning, whereas control participants exhibited fragmented recall and descriptive rather than explanatory accounts. The convergence of quantitative and qualitative evidence confirms that the Gagné-inspired POE sequence produced not merely more learning of the same kind, but qualitatively different learning characterized by conceptual coherence and transferable understanding.

Implications of the Study

This study demonstrates that constructivist learning principles can be effectively operationalized in resource-constrained Nigerian classrooms through instructional design that compensates for environmental limitations. The structured prediction phase externalized students' prior conceptions for examination despite large class sizes limiting individualized teacher interaction. The guided observation phase directed attention toward critical variables, compensating for limited laboratory access by ensuring students extracted maximum learning from each demonstration. The scaffolded explanation phase distributed cognitive processing across peer interactions, reducing dependence on ideal environmental conditions. These findings refine cognitive load theory's application in chemistry education, confirming that reducing extraneous cognitive load is essential for promoting the germane load necessary for schema construction when students coordinate macroscopic, microscopic, and symbolic representations. For practitioners, teachers must implement prediction activities requiring students to articulate reasoning underlying their predictions, incorporate structured observation protocols directing attention toward theoretically significant variables, and facilitate peer discussion with feedback that probes reasoning rather than simply confirming correct answers. Curriculum developers should design materials embedding Gagné's events within inquiry activities, including prediction sheets with reasoning prompts and explanation templates guiding students toward mechanistic accounts. Teacher training institutions must move beyond general advocacy for inquiry-based learning toward specific professional development in implementing structured inquiry sequences that promote conceptual restructuring rather than answer verification.

Limitations of the Research

The geographical specificity of this study limits broad generalisability, as participants came from ten schools in Sokoto State, Nigeria, representing northern Nigerian educational contexts that differ substantially from southern Nigeria and other sub-Saharan African countries. Replication across diverse contexts remains necessary before recommending widespread implementation. The reliance on self-reported perceptual data from student interviews, while triangulated with achievement measures, captures post-hoc reconstructions rather than real-time cognitive processing. Future research should employ think-aloud protocols during learning activities. The quasi-experimental

design, though appropriate for classroom research, cannot eliminate all potential confounding variables. Teacher effects, minimized through standardized protocols and fidelity monitoring, may still have influenced outcomes in ways not fully captured by the research design.

Recommendations for Future Research and Practice

Researchers should conduct replication studies examining the Gagné-inspired POE framework across diverse Nigerian regions and sub-Saharan African countries to identify contextual moderators of its effectiveness. Longitudinal designs must track students beyond four weeks to determine whether integrated knowledge structures support advanced chemistry learning. Comparative studies should investigate the framework's effectiveness for chemical bonding, thermodynamics, and electrochemistry topics. Professional development research must examine relationships between teacher training quality, implementation fidelity, and student outcomes. Studies should explore strategic integration of low-cost digital resources where infrastructure permits. Chemistry teachers should examine current POE practices for additional structure. Curriculum developers must revise materials incorporating Gagné's events. Teacher training institutions should integrate structured inquiry into programs. Policymakers should prioritize instructional quality funding over unreliable technological investments.

Reference

Abdugadar, F. A. (2026). Persistent misconceptions in acid-base chemistry among undergraduate science students: A qualitative investigation. *African Journal of Chemical Education, 14*(2), 112-128.

Adegoke, B. A., & Oladele, I. T. (2023). Digital simulation interventions for teaching titration concepts in Nigerian secondary schools: Challenges and opportunities. *African Journal of Educational Research, 41*(3), 245-262.

Adeyemi, S. B., & Oyedeji, O. A. (2022). Gagné-based instructional design in Nigerian secondary chemistry classrooms: Achievement and limitations. *Nigerian Journal of Science Education, 38*(1), 67-82.

Alkhabbas, F., Alzubi, M., & Al-Makhzoomy, A. K. (2022). Digital POE interfaces and particulate-level visualization in Jordanian university chemistry education. *International Journal of Science and Mathematics Education, 20*(4), 721-740.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.

Broman, K., Bernholt, S., & Parchmann, I. (2022). Context-based chemistry learning in German and Swedish classrooms: Interest, achievement, and conceptual understanding. *Journal of Research in Science Teaching, 59*(5), 823-848.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Dakic, M., Stojanovic, M., & Petrovic, N. (2025). POE implementations in European chemistry classrooms: A systematic review of structural limitations. *International Journal of Science Education, 47*(3), 401-418.

Dewi, A. D. M., Thohir, A., Nursobah, A., & Suntiah, R. (2025). Effectiveness of Technology-Enhanced Inquiry-Based Learning : A Meta-Analysis with Implications for Islamic Education. *Jurnal Pendidikan Islam*, *14*(2), 269–291. https://doi.org/https://doi.org/10.14421/jpi.2025.142.269291

Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11*(5), 481-490.

Jegstad, K. M. (2024). Inquiry-based chemistry education: a systematic review. *Studies in Science Education*, *60*(2), 251–313. https://doi.org/https://doi.org/10.1080/03057267.2023.2248436

Kaçar, T., Balım, A. G., & Evrekli, E. (2021). Inquiry-based learning in science education: A meta-analysis of effectiveness studies. *Journal of Research in Science Teaching, 58*(5), 684-710.

Kumar, S. (2025). Misconceptions in Acid-Base Chemistry: Diagnostic Assessment and Instructional Intervention. *International Journal of*

Applied and Behavioural Sciences (IJABS), *02*(02), 137–161. https://doi.org/https://doi.org/10.70388/ijabs250145

Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. Routledge.

Mirabueno, D. C., & Paderna, E. E. S. (2024). The Taxonomy of Predict-Observe- Explain (POE) as a Teaching Strategy and Thinking Process of Chemistry Stakeholders. *International Journal of Studies in Education and Science*, *5*(4), 390–403. https://doi.org/https://doi.org/10.46328/ijses.106

Molinar, R. V. (2026). Utilization of Gagne's Information Processing Model: Basis for Development of Instructional Materials in Teaching Teacher and the School Curriculum. *International Journal of Research and Innovation in Social Science*, *10*(26), 1689–1704.

Mueni, C., Rugano, P., & Mungai, J. (2023). Inquiry-based learning and students ' self-efficacy in Chemistry among secondary schools in Kenya. *Heliyon*, *9*(January), 0–9. https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e12672

Musa, G., Shehu, S. B., Ibrahim, S. A., Harazimi, A. C., & Ahmad, I. M. (2024). Effects of Gagne's Nine Events of Instruction on Biology Students Academic Performance in Tarauni, Kano State. *Ilorin Journal of Education (IJE)*, *44*(2), 92–105.

Nkosi, N. W., & Motlhabane, A. (2025). The Impact of Inquiry-Based Learning on Learners' Understanding of Chemical Change: A Quasi-Experimental Study. *Unesa Journal of Chemical EducationJournal of Chemical Education*, *14*(2), 127–138.

Özcan, G. E., & Uyanık, G. (2022). The effects of the "Predict-Observe-Explain (POE)" strategy on academic achievement, attitude and retention in science learning. *Journal of Pedagogical Research*, *6*(3), 103–111. https://doi.org/https://doi.org/10.33902/JPR.202215535

Piaget, J. (1968). *Six Psychological Studies*. Anita Tenzer (Trans.), New York: Vintage Books.

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.

Putriawati, H., & Chiani, S. H. (2025). Diagnostic assessment of acid-base misconceptions in Indonesian secondary schools. *Journal of Chemical Education, 102*(1), 45-58.

Rakhmalinda, F., Arifiani, R., & Puspawati, W. (2024). Analysis of Students ' Misconceptions on Acid-Base Topic Using the Two-Tier Diagnostic Test. *Journal of Computers for Science and Mathematics Learning*, *1*(1), 27–43.

Sabiru, M. (2013). Effects of Gagné's instructional model on students' academic achievement in chemistry. *Journal of Educational Research and Development, 8*(2), 112-124.

Sari, S. D. C., Ardi, T., Parameswary, C. S., & Widiastuti, N. L. (2025). Identifying Students' Understanding Levels and Sources of Misconceptions in Acid–Base Concepts through a Five-Tier Diagnostic Test. *EduChemia Jurnal Kimia Dan Pendidikan*, *10*(1), 163–174. https://doi.org/10.62870/educhemia.v10i1.35563

Shaafi, A. M., Rahman, N. A., & Yusoff, S. (2025). Pre-service science teachers' understanding of acid-base concepts: A multi-country diagnostic study. *International Journal of Science and Mathematics Education, 23*(2), 289-306.

Shaafi, N. F., Yusof, M. M. M., Ellianawati, E., Subali, B., & Raji'e, M. H. H. (2025). Investigating misconceptions about acids and bases among pre-service science teachers. *Journal of Education and Learning*, *19*(1), 460–477. https://doi.org/10.11591/edulearn.v19i1.21803

Shehu, S., Ukor, D. D., Omiola, M. A., & Ibrahim, S. T. (2025). Effect of Gagne's-Nine-Events-of-Instruction on Academic Performance in Genetics among Biology Students in Kumbotso Zonal Education Directorate, Kano State, Nigeria. *Rima International Journal of Education (RIJE)*, *4*(2), 101–113.

Sweller, J. (1988). Cognitive Load During Problem Solving : Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/Doi: 10.1016/0364-0213(88)90023-7

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 10*(2), 159-169.

Twizeyimana, E., & Shyiramunda, T. (2026). The Effect of Inquiry-Based Chemistry Laboratory Instruction on Students' Conceptual Understanding: A Meta- Analysis. *Sage Open*, *16*(3), 1–16. https://doi.org/10.1177/21582440261466348

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. MA: Harvard University Press.

Wang, L., Chen, Y., & Zhang, H. (2023). Applying Gagné's instructional events in Chinese secondary biology classrooms. *Journal of Biological Education, 57*(4), 389-404.

Zakiyah, I. (2019). The Effectiveness of Predict-Observe-Explain Strategy to Reduce Misconception in Thermochemistry. *Advances in Computer Science Research*, *95*(1), 139–142.