23 January 2026

Revisiting a Common Cobalt Chloride Equilibrium Experiment with Generative Artificial Intelligence

Devin M Mulvey[1], Katsu Ogawa[1,2], Alyssa V B Santos[3], Henri C Santos[4], Patrick W Schneider[1], Scott Simpson[1]

1. Department of Chemistry St. Bonaventure University
2. Department of Natural Science University of Jamestown
3. Department of Chemistry Virginia Tech
4. Lirio, LLC

Abstract

We have developed and implemented a first-year undergraduate laboratory experiment at St. Bonaventure University that augments a well-established thermodynamics lab on the spectrophotometric determination of the equilibrium of cobalt(II) chloride hexahydrate with generative artificial intelligence (GenAI). Students utilized free ChatGPT accounts as a support tool in their analysis of laboratory data and iteratively refined their answers to chemical questions. Overall, the inclusion of GenAI in student workflow did not appear to negatively impact student performance, but there were instances where it struggled to produce correct answers for tasks. Descriptive analysis of student responses, chat logs, and instructor discussions suggests that students were resilient to incorrect GenAI output. This laboratory experiment is designed such that it can be easily adopted by faculty with an interest in exploring the use of GenAI in chemistry curricula.

Keywords

First-Year Undergraduate, Web-Based Learning Communication, Writing Multimedia-Based Learning, General Physical Chemistry, Internet, Interdisciplinary, Multidisciplinary Laboratory Instruction, Calorimetry, Thermochemistry, Equilibrium, UV-Vis Spectroscopy, Generative Artificial Intelligence

Revisiting a Common Cobalt Chloride Equilibrium Experiment with Generative Artificial Intelligence

Devin M. Mulvey*,[a], Katsu Ogawa[a,b], Alyssa V. B. Santos[c], Henri C. Santos[d], Patrick W. Schneider [a], and Scott Simpson [a]

[a]Department of Chemistry, St. Bonaventure University, St Bonaventure, NY, 14778, USA

[b]Department of Natural Science, University of Jamestown, Jamestown, ND, 58401, USA

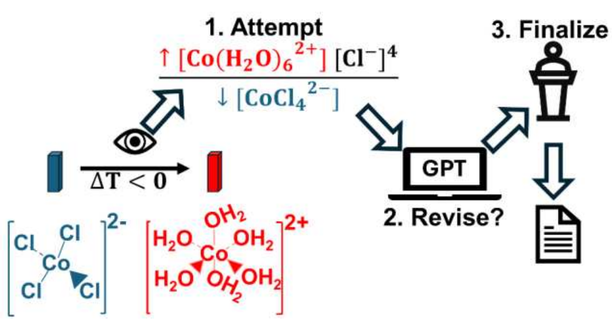
[c]Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA

[d]Lirio, LLC, Knoxville, TN, USA

ABSTRACT

We have developed and implemented a first-year undergraduate laboratory experiment at St. Bonaventure University that augments a well-established thermodynamics lab on the spectrophotometric determination of the equilibrium of cobalt(II) chloride hexahydrate with generative artificial intelligence (GenAI). Students utilized free ChatGPT accounts as a support tool in their analysis of laboratory data and iteratively refined their answers to chemical questions. Overall, the inclusion of GenAI in student workflow did not appear to negatively impact student performance, but there were instances where it struggled to produce correct answers for tasks. Descriptive analysis of student responses, chat logs, and instructor discussions suggests that students were resilient to incorrect GenAI output. This laboratory experiment is designed such that it can be easily adopted by faculty with an interest in exploring the use of GenAI in chemistry curricula.

GRAPHICAL ABSTRACT



KEYWORDS

First-Year Undergraduate / General Physical Chemistry

Interdisciplinary / Multidisciplinary Laboratory Instruction

Internet / Web-Based Learning Communication / Writing Multimedia-Based Learning

Calorimetry / Thermochemistry

Equilibrium

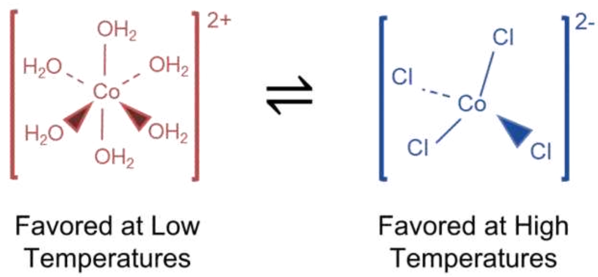
UV-Vis Spectroscopy

Generative Artificial Intelligence

INTRODUCTION

The temperature and ligand concentration dependence of equilibrating cobalt(II) complexes is a well-trodden undergraduate general chemistry experiment with multiple iterations of the procedure resurfacing in the literature for decades.[1-5] Herein, we consider the complex pair cobalt(II) hexahydrate ($[Co(H_2O)_6]^{2+}$) and cobalt(II) tetrachloride ($[CoCl_4]^{2-}$). Schematic 1 shows the reaction of interest and the Lewis structures of the two complexes.

$$Co(H_2O)_6^{2+}(aq) + Cl_4^{-}(aq) \rightleftharpoons CoCl_4^{2-}(aq) + 6H_2O(l)$$



Schematic 1. Cobalt(II) hexahydrate and cobalt(II) tetrachloride equilibrium reaction in an aqueous environment. Coloration of the Lewis structures approximates the hue of the complex as it appears in water.

The popularity of this reaction is evident given the distinctly colored species (pink and blue respectively) provide a visual confirmation of the dominant complex as students manipulate reaction conditions. Additionally, the λ_{max} of these complexes (~520 nm for $[Co(H_2O)_6]^{2+}$ and ~690 nm for $[CoCl_4]^{2-}$) are well separated making it feasible for students to perform spectrophotometry, determine the equilibrium constant, and the thermodynamic constants ΔH°, ΔS°, and ΔG°. Such a well-established experiment is fertile ground for augmentation with generative artificial intelligence (GenAI).

GenAI tools are becoming increasingly pervasive in many areas of daily life, including academia. Surveys indicate that the majority of college students use generative AI tools for coursework, often on a weekly basis, and would like universities to provide training on their effective, professional, and ethical use.[6,7] This desire is not unwarranted as McKinsey’s 2025 report revealed that adoption of AI by businesses continues to grow with 88% of

1993 participants reporting regular AI use in at least one business function[8]. These findings underscore the desire of and need for students to develop proficiency with these tools.

A rapidly growing body of literature reflects the academic community's experimentation with and adoption of these tools. Recent studies incorporate GenAI into development of core scientific problem-solving competencies like writing, literature review and critique,[9-14] utilize GenAI as a tool for data generation and analysis,[15-16] and employ GenAI for personalized study.[17-18] Parallel studies examine student use of GenAI, AI–student failure modes, and instructional or technological solutions such as retrieval-augmented generative AI (RAGAI).[19-25]

Despite a growing body of lecture-based GenAI activities, few laboratory experiments incorporate GenAI, particularly at the first-year undergraduate chemistry level. This gap in literature and desire to incorporate formal opportunities to use GenAI in chemistry led this work. Herein, we augment a well-established general chemistry laboratory experiment on the thermodynamics and equilibrium of cobalt complex pair with GenAI. Additionally, we explore how students iterate their answers when given the opportunity to use GenAI with observations on student-AI interactions throughout the process.

EXPERIMENTAL OVERVIEW

The scientific learning objectives of this experiment are well established by the work that was adapted for the wet-lab portion of the experiment[3]. The student learning objectives (SLOs) we are measuring focus on integration of GenAI. The SLOs are as follows:

1. Interpret and analyze experimental equilibrium and thermodynamic data obtained from UV-Vis spectroscopy.
2. Evaluate generative AI output related to chemical calculations and symbolic expressions for accuracy, consistency, and alignment with experimental evidence.
3. Refine and justify chemical conclusions by integrating experimental data, AI feedback, peer discussion, and instructor guidance.

The experiment was conducted during a 3-hour laboratory period at St. Bonaventure University (SBU) across three laboratory sections, each enrolling up to 17 students, and a single instructor per section. Students worked in groups of two to three to collect experimental data and were permitted to confer with their instructor during the experimental portion of the lab. A total of 49 students participated in the experiment, most in their second semester of first-year general chemistry; subsequent results include only 35 students due to incomplete submission of required work. Students belonged to variety of majors (Biochemistry, Biology, Chemistry, Environmental Science, Health Science, and Physics). The students had covered chemical equilibrium in lecture by this point, but not its connection to thermodynamic quantities ($\Delta H°$, $\Delta S°$, and $\Delta G°$).

The week before the experiment digital training materials were distributed using SBU's learning management system. The materials included four pre-recorded videos which covered: how to setup a ChatGPT account; a top-level introduction to GenAI, LLMs, neural networks, RAGAI, and safety considerations; an introduction to prompt engineering using AI for Education's Five S's framework[26]; and a pre-lab lecture on the thermodynamic theory and its connection to the physical lab experiment. A 10-question pre-lab quiz was distributed as well, which was meant to prime students for the upcoming experiment and gauge their familiarity with GenAI, lab procedure, and underlying thermochemical theory. All of these materials are provided in the SI.

Students were required to bring personal laptops and a USB stick. Internet access was available to students during the entirety of the laboratory period. The experimental portion of the exercise consisted of three sections:

- Section A - Preparation of Sample Solution: In this section, students prepared a cobalt(II) chloride hexahydrate sample solution in a stock HCl solution.
- Section B – Collecting Full Absorbance Spectrum of Solutions: In this section, students used a SpectroVis Plus spectrophotometer from Vernier[27] or Red Tide USB650 spectrometer from Ocean Optics[28] paired with a LabQuest 2[29] to measure the full absorbance spectrum of their sample and a stock $[CoCl_4]^{2-}$ complex solution. Students plot absorbance versus wavelength to identify λ_{max} of the $[CoCl_4]^{2-}$ complex.
- Section C – Measuring Absorbance and Temperature Over Time: In this section, students measure the change in concentration of $[CoCl_4]^{2-}$ as a warmed portion of their sample solution cools off using their spectrophotometer, LabQuest 2, and a Vernier temperature probe[30].

Post-Lab Analysis

After the experiment, students were given seven days to complete a lab report consisting of three parts (I–III). Part I included six computational questions requiring analysis of experimental data and use of ChatGPT and was graded for accuracy. Part II contained four open-ended questions probing conceptual understanding of equilibrium, thermodynamics, and experimental error using ChatGPT and a custom RAGAI, and was also graded for accuracy. Part III comprised five open-ended questions assessing student perceptions of the activity, AI use, and related safety considerations, and was graded for completion. A grading guide is provided in the SI Instructor Notes. For this work, we focus on the student responses in Part I. Parts II and III delve deeper into student perceptions rather than quantifiable correct and incorrect answers. As such, these will be analyzed in future work.

The questions in Part I were structured such that students attempted the same task three times. Table 1 provides a sample question structure.

Table 1. Sample Table of Format for Student Responses to Pre- and Post-Assessment Questions

WITHOUT AI (1 pt):	**WITH AI (1 pt):**
Graded on completion	Graded on completion
Final Answer (3 pts):	
Graded for accuracy	

A student's first attempt was made before asking ChatGPT (Phase *B*: *B*efore AI) and on their second attempt they were instructed to ask the same question of ChatGPT using their first attempt as a starting place for discussion and write the resulting output (Phase *A*: *A*fter AI). The deidentified chat logs for part I questions are available in the SI. Before students started prompting ChatGPT, they were instructed to upload an abridged copy of the lab procedure, contained in the SI. These first two attempts were graded upon completion and students were not restricted from consulting their peers. They were not permitted to communicate with the instructor as it was necessary to assess whether students could be successfully redirected by GenAI without instructor intervention. On the third and final attempt the students were allowed to consult their instructor (Phase *F*: *F*inal Answer).

Table 2 lists the symbols used throughout this work to refer to a particular assessment task, a summary of what it entails, and the framework used to bin it as correct, partially correct, incorrect, or consistent in subsequent analyses performed.

Table 2. Assessment Tasks, the Symbols Used to Represent them, and the Framework Employed for Binning Student Answers as Correct, Partially Correct, Incorrect, and Consistent

Symbol Used	Assessment Task	Correct* Criteria	Partially Correct Criteria	Incorrect Criteria	Consistent Criteria
ε	Calculate molar absorptivity of cobalt(II) tetrachloride ($[CoCl_4]^{2-}$) complex using Beer's law and the spectrophotometric data collected from a standard solution.	• Significant Figures • Units	• Magnitude	Answer fails to meet Partially Correct criteria.	
K_{sym}	Write the symbolic equilibrium expression for the direction they ran the reaction in lab (cooling solution off).	• Notation	• Ratio • Chemical Species • Exponent magnitude		
$\Delta H°$	Extract the standard change in enthalpy of reaction in kJ/mol from a plot of ln(K) vs. 1/T and the Van't Hoff Equation.	• Significant Figures • Units	• Sign • Magnitude		Technically incorrect due to a mistake(s) made in a previous assessment task(s), but analysis of the incorrect starting values is correct for the current task.
$\Delta S°$	Extract the standard change in entropy of reaction in J/(mol·K) from a plot of ln(K) vs. 1/T and the Van't Hoff Equation.				
$\Delta G°$	Calculate the standard change in Gibbs free energy of reaction in kJ/mol using ΔH°, ΔS°, and T=25℃.				
K_{num}	Use the Arrhenius equation and ΔG° from the previous step to calculate a numerical value for the equilibrium constant.		• Magnitude		

*The "Correct" column shows additional criteria to meet to get full credit. Meaning for a response to be binned as "Correct" it must meet *all* the criteria in "Partially Correct" *and* the criteria in the "Correct."

Before proceeding it is worth clarifying what constitutes a correct, partially correct, incorrect, and consistent binning. For task ε, if an answer has the correct magnitude, the correct number of significant figures, and units of $L \cdot mol^{-1} \cdot cm^{-1}$ the answer is binned as correct. If it only has the correct magnitude, it is binned as partially correct. For K_{sym} the ratio and identity of chemical species, exponents on chemical species, and standard notation for chemical species must all be correct for an answer to receive a correct binning, like so: $\frac{[Co(H_2O)_6^{2+}][Cl^-]^4}{[CoCl_4^{2-}]}$. The partially correct bin is reserved for answers that make slight mistakes in notation like forgetting to include a charge. There is no consistent bin for ε and K_{sym} because they do not use the answer(s) from preceding assessment task(s). For assessment tasks ΔH°, ΔS°, ΔG° the answer must have the correct magnitude, number of significant figures, units (kJ/mol, J/mol·K, and kJ/mol respectively), and sign (-, +, and - respectively) to be binned as correct. If an answer only gets sign and magnitude correct it is binned as partially correct. Task K_{num} has the same binning criteria as task ε. Answers binned as consistent are technically incorrect for a student's lab data but are consistent with the values of previous assessment tasks. If the answer does not meet any of the listed criteria, it is binned as incorrect.

EXPERIMENTAL DESIGN LIMITATIONS

As a final note regarding experimental design, we would like to point out limitations of this study. By grading the first two attempts before consulting an instructor on completion, we attempted to disincentivize students from blatantly copying GenAI output and their peers work as their own. We assume that they followed these instructions. We did not check all student chat logs exhaustively. Rather, we utilized student responses in part I of the lab report and submitted excel workbooks to bin answers. Those results allowed us to identify specific chat logs or general trends worth discussing.

HAZARDS AND SAFETY PRECAUTIONS

The hazards associated with this experiment arise from the irritant, corrosive, mutagenic, carcinogenic, and aquatic toxicity characteristics of cobalt(II) chloride hexahydrate as well as the corrosive characteristics of hydrochloric acid. Work should be performed in a fume hood to reduce exposure to hydrogen chloride gas. Students should wear goggles, gloves, and protective clothing. Cobalt solutions should be disposed of as a heavy metal salt containing solution. Heated solutions should be approached with caution and not heated in a sealed container.

INSTITUTIONAL REVIEW BOARD

The Institutional Review Board (IRB) received the protocol proposal on February 14, 2025, assigned it protocol ID #727, and approved it as an exempt educational research study on February 21, 2025; documentation and details are provided in the SI Instructor Notes.

RESULTS & DISCUSSION

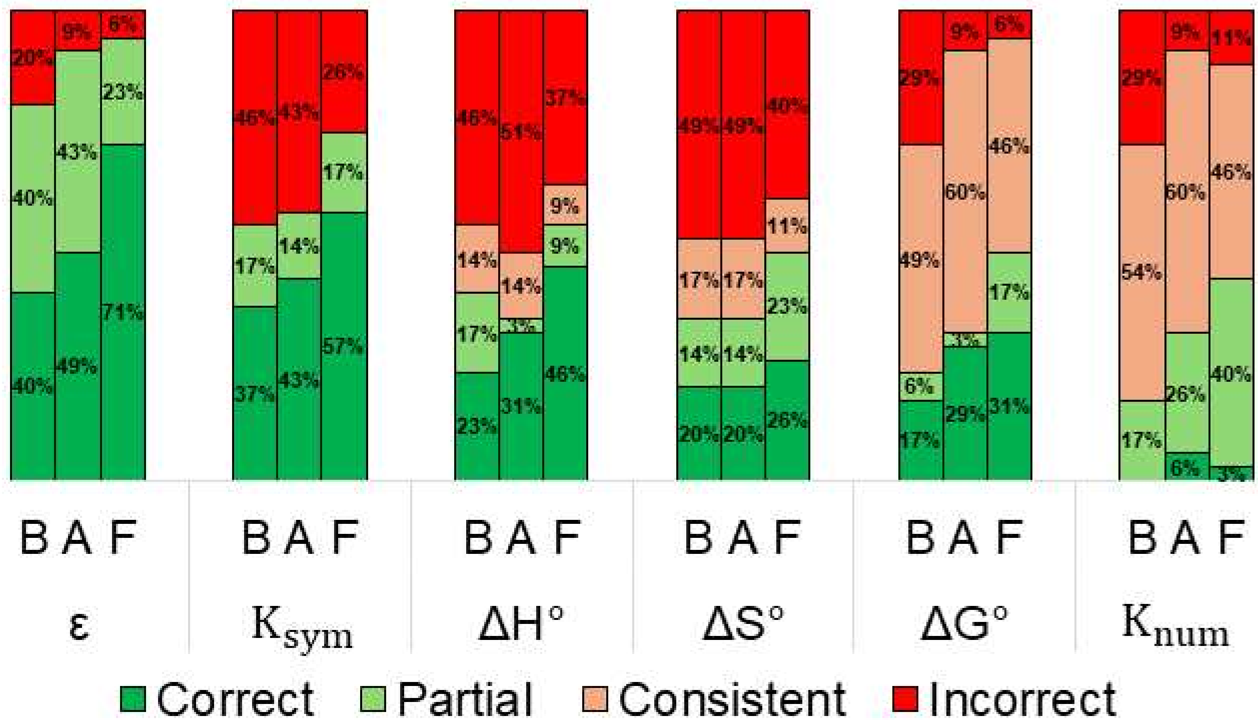


Figure 1. Stacked bar plot showing the percentage of correct, partially correct, consistent, and incorrect responses to assessment tasks before consulting GenAI (*B*), after consulting GenAI (*A*), and after being given the opportunity to confer with lab Instructor (*F*). Note that the % in the plot are rounded to the nearest whole number, so they may not add up exactly to 100%. Sample size of N=35 students.

Figure 1 shows the percentage of correct, partially correct, consistent, and incorrect responses to assessment tasks Each column represents a different stage in developing a final answer: Phase *B* shows students' answers before consulting ChatGPT, Phase *A* shows the responses generated by ChatGPT, and Phase *F* shows students' final answers after being allowed to discuss the task with their lab instructor. Comparing phase *B* and phase *A* of each assessment task reveals that the number of correct and partially correct responses increased (≤ 14.3%) for assessment tasks ε, $\Delta G°$, and K_{num}, did no change for task $\Delta S°$, and decreased (5.7%) for $\Delta H°$. If one incorporates the answers binned as consistent, there was an additional increase of 20% for tasks $\Delta G°$, and K_{num}. *It is important to note* that the increase in correct answers when going from phase *B* to phase *A* may be the result of ChatGPT responding with a correct answer, students consulting their classmates, or a combination of both. Across all assessment tasks, final student answers were correct more often than the previous two attempts. The increase in correct answers when going from phase *A* to *F* may be the result of student-instructor interactions and/or discussions with classmates.

The general upward trend in correct students answers from phase *B* to *F* indicates that the inclusion of ChatGPT in student workflow does not appear to negatively impact overall student performance. However, comparing phases *B* and *A* of tasks $\mathrm{K_{sym}}$, ΔH°, and ΔS° indicates that there are instances where ChatGPT produced incorrect answers. We use Sankey flow diagrams[31-34] to illustrate how students transition between correct and incorrect responses across the three assessment phases. In these diagrams, nodes represent student groups and connecting arcs represent transitions between them, with node and arc heights proportional to the number of students involved.

Because this study involved a small sample size (35 students), which was further divided into smaller subgroups, the results should not be considered statistically significant or representative of the general student population. Instead, Figures 2 and 3 are used as descriptive tools to support discussion of how GenAI was integrated into student workflows and to highlight observations that may be useful to instructors interested in adopting this laboratory experiment.

The first two assessment tasks ε and $\mathrm{K_{sym}}$ serve as illustrative examples with easily observed and explained phenomena. Figure 2 shows the number of correct (C) and incorrect (I) responses to the assessment tasks ε and $\mathrm{K_{sym}}$ at each step along the process of arriving at a final answer.

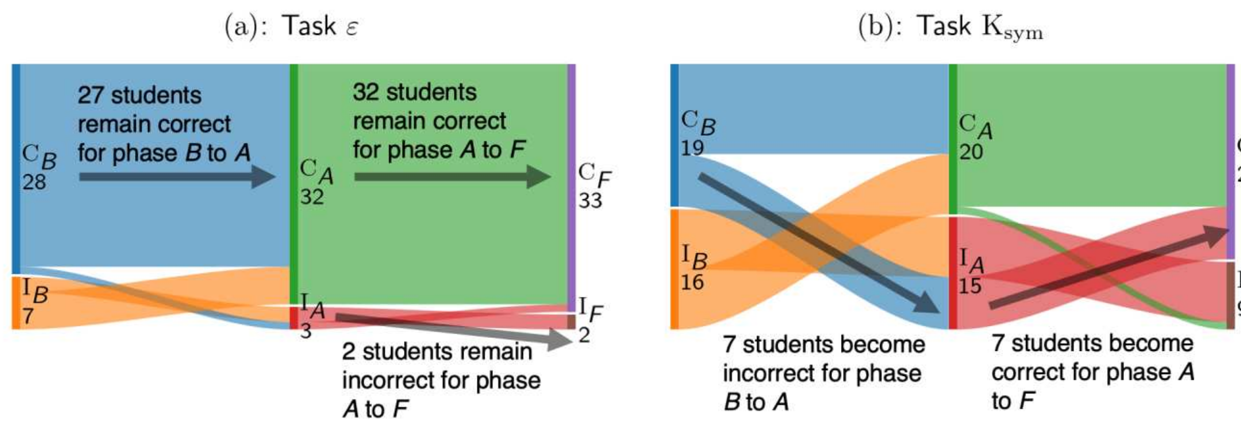


Figure 2. Sankey flow diagrams of student populations moving between correct (C) and incorrect (I) response nodes, before consulting ChatGPT (*B*), after consulting ChatGPT (*A*), and after being given the opportunity to confer with lab Instructor (*F*) for assessment tasks ε (a) and $\mathrm{K_{sym}}$ (b). The numerical value below each tag for a node represents the number of students the node contains. Nodes represent student groups and connecting arcs represent transitions between them, with node and arc heights proportional to the number of students involved. Note that correct nodes (C) include both fully and partially correct answers. Overlayed on top of the flow diagram are text-based explanations of representative arcs and corresponding arrows showing the direction in which students "flow between nodes". Sample size of N=35 students.

For assessment task ε, 28 students started with correct answers, 27 retained that binning with 5 joining them after writing the output of ChatGPT, and 32 maintained it in their final answer with the addition of 1 student.

Interestingly, only one student who started with a correct answer and has an incorrect answer in phase *A* (i.e. there is a narrow band traveling from C_B to I_A). Conversely, 5 students who started with an incorrect answer in phase *B* wrote a correct answer in phase *A* which should represent the output of ChatGPT for the same assessment task. In looking at the chat logs, it was found four of those students were corrected by ChatGPT via identification of common mistakes like using the wavelength of maximum absorption (λ_{max}) rather than the maximum absorption (A) in the Beer-Lambert law.

It is worth noting that the ChatGPT did struggle to provide units often. It appeared to attempt using LaTex formatting in its outputs, but this would not render properly, with one student pushing for ChatGPT to "write this properly". In scenarios where units are needed and the AI failed to produce easily read text it is possible that students who wrote the correct units in their lab reports consulted their peers.

This is a demonstration of a best-case scenario: In the case where students start off with a correct response (C_B), the vast majority remain correct throughout the process; In the case where students start off incorrect (I_B), the output of ChatGPT and/or peer interactions lead to corrections. Above all, students are course corrected and retained with most inter-node travel being upward (I to C) and about 94% of students having a correct final answer.

For assessment task $\mathrm{K_{sym}}$ there is notably more movement between the nodes at all phases. In stark contrast to task ε, nearly half the students (16) started with incorrect equilibrium expressions. Of those 16 students, one half wrote an incorrect answer in phase *A,* while the other half wrote a correct answer. Notably, 7 of the 19 students that started with correct answers in phase *B* wrote an incorrect answer in phase *A*. The observed decrease in correct responses was due in phase *A*, in part, to ChatGPT being unable to determine which direction the reaction was performed in. The reaction was presented in both directions (endo- and exothermic) in the lab report, but for this assessment task students were explicitly asked to write $\mathrm{K_{sym}}$ when the reaction is run in the exothermic direction. Despite this, some student interactions with ChatGPT led to inverted expressions for the equilibrium expression:

$$\mathrm{K_{sym}^{-1}} = \frac{[\mathrm{CoCl_4^{2-}}]}{[\mathrm{Co(H_2O)_6^{2+}}][\mathrm{Cl^-}]^4}$$

In checking the chat logs, 14 of the 15 students with incorrect answers in phase *A* either did not specify a direction the reaction was run in (cooling or heating), provided the incorrect direction, or provided incomplete information in their prompt to ChatGPT. Conversely, all 12 students with correct answers for both phases *B* and *A* and 6 of the 8 students whose answers were corrected in phase *A*, either specified the direction the reaction was run in or the specific section of text which shows the exothermic version. It is clear that a lack of context in prompting has a profound effect on ChatGPT being able to produce $\mathrm{K_{sym}}$ correctly**.** Interestingly, 6 of the 14 students with incorrect answers for $\mathrm{K_{sym}}$ in phase *A* either argued with the ChatGPT or tried obtaining the equilibrium expression multiple times. This strategy worked for 4 of the 6 students, as evidenced by their adoption of a correct final answer.

Figure 2b paints a clear picture of when student-GenAI interactions can go awry. When students do not practice prompt engineering skills, like providing sufficient context, GenAI can struggle to assist them with simple tasks like writing an equilibrium expression. However, through peer discussions, instructor intervention, and the rejection of incorrect GenAI output, the same number of students who adopted incorrect answers in phase *A* switched to a correct final answer in phase *F*. Although the process was not as smooth as assessment task ε, the majority of expressions for $\mathrm{K_{sym}}$ (74%) were correct in the final phase, demonstrating a resilience to GenAI failure in student workflows.

All tasks following $\mathrm{K_{sym}}$ are coupled which led to scenarios where students had generated incorrect values for an earlier assessment task, but all subsequent analysis with that bad value was correct. This necessitated the creation of a “consistent” bin in Figure 1 and consequently a “consistent” node (C/I) must be included in the Sankey flow diagrams for assessment tasks ($\Delta\mathrm{H}°$, $\Delta S°$, $\Delta\mathrm{G}°$, and $\mathrm{K_{num}}$). Figure 3 aggregates the flow diagrams for all four tasks.

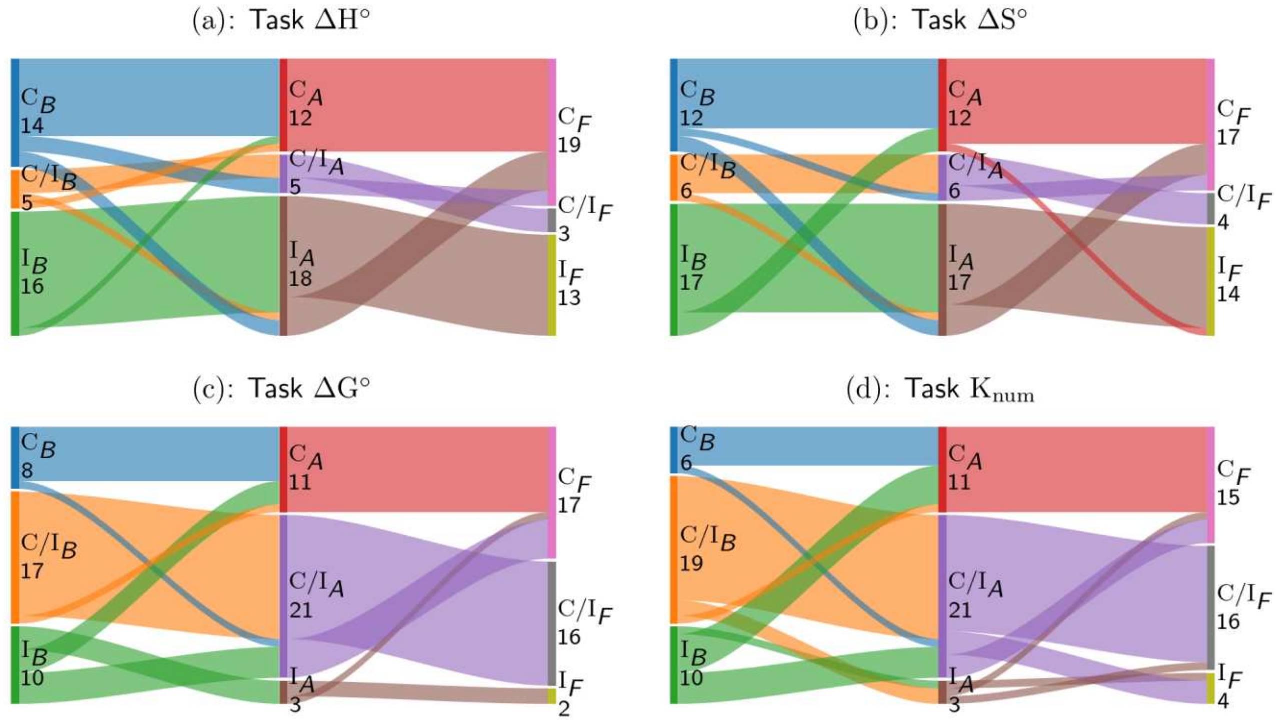


Figure 3. Sankey flow diagrams of student populations moving between correct, consistent, and incorrect response nodes, C, C/I, and I respectively, before consulting ChatGPT (*B*), after consulting ChatGPT (*A*), and after being given the opportunity to confer with lab Instructor (*F*) for assessment tasks $\Delta\mathrm{H}°$ (a), $\Delta S°$ (b), $\Delta\mathrm{G}°$ (c), and $\mathrm{K_{num}}$ (d). The numerical value below each tag for a node represents the number of students the node contains. Nodes represent student groups and connecting arcs represent transitions between them, with node and arc heights proportional to the number of students involved. Note that correct nodes (C) include both fully and partially correct answers. Sample size of N=35 students.

The sheer complexity of movement between nodes in these diagrams makes it prohibitive to provide complete explanations of what is causing student answers to change. The narrow arcs in the above plots represent the movement of *very small* sub-populations of students. For instance, the inter-node travel from phase *B* to phase *A* of assessment task $\Delta\mathrm{H}°$ contains eight unique arcs connecting six nodes. We do not attempt to assign causality to individual changes, as doing so would be speculative, and instead focus on discussing the general trends observed in the plots.

For the ΔH° and ΔS° assessment tasks, Figures 3(a) and 3(b), student responses are polarized across all three phases, with most answers being either correct or incorrect and fewer falling into the consistent bin. For task $\Delta\mathrm{H}°$ ($\Delta S°$),10 (9) students start with correct answers in phase *B* and retain that binning in phase *A*, while 15 (14) have incorrect answers for the first two phases. For both tasks, most inter-node movement between the *A* and *F* phases is upward leading to more correct/consistent final answers than incorrect. This polarization of student answers is understandable considering the number and complexity of steps students must take to arrive at $\Delta H°$ and $\Delta S°$, namely students must perform the following in an Excel workbook:

1. Convert temperature data to K^{-1} from ℃.
2. Calculate $[CoCl_4^{2-}]$, $[Cl^-]$, and $[Co(H_2O)_6^{2+}]$ from solution concentration and absorbance data they collected in lab.
3. Calculate the value of the equilibrium constant K at each temperature from the concentrations in 2.
4. Plot the natural log of K vs. 1/T and fit the data with a linear trend.

The sheer number of steps and potential points of failure in determining $\Delta H°$ and $\Delta S°$ undoubtedly contributed to the number of incorrect responses. We did not prevent students from uploading their plots or Excel workbooks to ChatGPT during phase *A* of extracting $\Delta H°$ and $\Delta S°$ from their spectrophotometric data. However, only 7 students shared some portion like a sheet or figure from their workbook with ChatGPT. It is unclear whether this choice was intentional or due to unclear communication of lab guidelines, as some students reported uncertainty about whether ChatGPT could be used to assist with development of formula and/or review their Excel files. While there is no guarantee that students sharing their raw data with ChatGPT would lead to better outcomes, but future iterations of this work should more clearly identify where and when students can utilize GenAI.

For assessment tasks $\Delta G°$ and K_{num}, Figures 3(c) and 3(d), we see a different pattern appear. Most students start with answers that are technically incorrect for their lab data but are consistent with the values of $\Delta H°$ and $\Delta S°$ they calculated in the preceding steps (C/I_B = 17 for $\Delta G°$ and C/I_B = 19 for K_{num}). Of the remaining students, more start with incorrect answers in phase *B* of both assessment tasks. In phases *A* and *F* there is general upward trend for $\Delta G°$ with the total number of correct and consistent responses increasing from $C_B + C/I_B$: 25 to $C_A + C/I_A$: 32 to $C_F + C/I_F$: 33. For K_{num} the outcome is largely the same except there is one less student in the C_F and C/I_F bins relative to phase *A*. The decrease in incorrect answers for tasks $\Delta G°$ and K_{num} when compared to $\Delta H°$ and $\Delta S°$ is likely related to computational simplicity and low complexity of determining $\Delta G°$ and K_{num}.

CONCLUSIONS

We developed a first-year undergraduate laboratory experiment that augments a well-established cobalt complex equilibrium thermodynamics lab with generative artificial intelligence (GenAI). Designed as a drop-in addition to general chemistry curricula, the experiment uses free ChatGPT accounts to support data analysis and refinement of chemical reasoning.

Incorporating GenAI did not negatively affect student performance, although ChatGPT occasionally struggled with tasks such as writing units and symbolic equilibrium expressions. Analysis of student responses and chat logs indicates that students were resilient to these shortcomings through peer discussion and instructor guidance.

Because of the small sample size (35 students) and collaborative assessment design, the results are not broadly generalizable; nevertheless, this work motivates future studies on student–AI interactions, instructor-mediated discussion, and the use of tools such as RAGAI and prompt-engineering strategies to improve AI-supported instruction.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dmulvey@sbu.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

SS and DMM acknowledge the National Science Foundation (Award #2142874), Camille & Henry Dreyfus Foundation (Award # TH-23-033), and the Leo E. Keenan Faculty Endowment for support of this research. The authors acknowledge Adam Brown from St. Bonaventure University for providing feedback on the structure of the project and assistance with the IRB filing. The content is solely the responsibility of the authors and does not necessarily represent the views of Lirio (HCS affiliation).

REFERENCES

1. Ophardt, C. E. Cobalt complexes in equilibrium. *J. Chem. Educ.* **1980**, *57*, 453.

2. DeGrand, M. J.; Abrams, M. L.; Jenkins, J. L.; Welch, L. E. Gibbs Energy Changes during Cobalt Complexation: A Thermodynamics Experiment for the General Chemistry Laboratory. *J. Chem. Educ.* **2011**, *88*, 634–636.

3. Antony, E.; Muccianti, C.; Vogel, T. An Experimental Determination of Thermodynamic Values. *J. Chem. Educ.* **2012**, *89*, 533–535.

4. Ren, J.; Lin, T.; Sprague, L. W.; Peng, I.; Wang, L.-Q. Exploring Chemical Equilibrium for Alcohol-Based Cobalt Complexation through Visualization of Color Change and UV–vis Spectroscopy. *J. Chem. Educ.* **2020**, *97*, 509–516.

5. Zacher, B.; Ferrante, S.; Romero, R. Thermochromism: Theory Directed Design of a Thermochromic Thermometer. *J. Chem. Educ.* **2023**, *100*, 796–802.

6. Flaherty, C. How AI Is Changing—Not 'Killing'—College. 2025; https://www.insidehighered.com/news/students/academics/2025/08/29/survey-college-students-views-ai, (accessed Jan 10, 2025).

7. Digital Education Council. Global AI Student Survey 2024. 2024; https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024, (accessed Jan 10, 2025).

8. Singla, A.; Sukharevsky, A.; Hall, B.; Yee, L.; Chui, M.; Balakrishnan, T. The State of AI: Global Survey 2025 — McKinsey. 2025; https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/, (accessed Jan 10, 2025).

9. Yuriev, E.; Burton, M. G.; Knott, D.; Terrill, A. E.; Jackson, N. R. C.; Chen, S. Integrating Generative AI in Chemistry Education: Enhancing Career-Ready Writing Skills in Pharmaceutical Science. *J. Chem. Educ.* **2025**, *102*, 1991–2001.

10. Reddy, M. R.; Walter, N. G.; Sevryugina, Y. V. Implementation and Evaluation of a ChatGPT-Assisted Special Topics Writing Assignment in Biochemistry. *J. Chem. Educ.* **2024**, *101*, 2740–2748.

11. Varju, B. R.; Noman, M. E.; Dicks, A. P.; Seferos, D. S. Incorporating ChatGPT into a Simulated Peer-Review Assignment in an Upper-Level Materials Chemistry Course. *J. Chem. Educ.* **2025**, ASAP Article. https://doi.org/10.1021/acs.jchemed.5c00790.

12. Pence, H. E.; Hightower, G.; Forlenza, J.; Leonard, K.; McLellan, A.; Suero, A.; Amoah, B.; Mbow, M.; Borner, S.; Castillo, A. et al. Using Generative AI Systems for Critical Thinking Engagement in an Advanced Chemistry Course: A Case Study. *J. Chem. Educ.* **2024**, *101*, 3789–3794.

13. Guo, Y.; Lee, D. Leveraging ChatGPT for Enhancing Critical Thinking Skills. *J. Chem. Educ.* **2023**, *100*, 4876–4883.

14. Exintaris, B.; Karunaratne, N.; Yuriev, E. Metacognition and Critical Thinking: Using ChatGPT-Generated Responses as Prompts for Critique in a Problem-Solving Workshop (SMARTCHEMPer). *J. Chem. Educ.* **2023**, *100*, 2972–2980.

15. García-Morales, M. Automating Excel Tasks with Generative AI: Supporting Innovation in Mass Transfer Separations Education. *J. Chem. Educ.* **2025**, ASAP Article. https://doi.org/10.1021/acs.jchemed.5c01531.

16. Hare, P. M. Coding with AI in the Physical Chemistry Laboratory. *J. Chem. Educ.* **2024**, *101*, 3869–3874.

17. Thway, M.; Recatala-Gomez, J.; Lim, F. S.; Hippalgaonkar, K.; Ng, L. W. T. Harnessing GenAI for Higher Education: A Study of a Retrieval Augmented Generation Chatbot's Impact on Learning. *J. Chem. Educ.* **2025**, *102*, 3849–3857.

18. Vincent, M. A.; Lear, B. J. Using Exam Preparation and Reflection to Introduce Artificial Intelligence Tools in Honors General Chemistry. *J. Chem. Educ.* **2025**, *102*, 4470–4478.

19. Tassoti, S. Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education. *J. Chem. Educ.* **2024**, *101*, 2475–2482.

20. Sigot, M.; Tassoti, S. A Missed Opportunity for No-Code Chatbots? Current Challenges in Publicly Available Chemistry GPTs. *J. Chem. Educ.* **2025**, *102*, 2151–2159.

21. Sigot, M.; Tassoti, S. An Investigation of Change in Prompting Strategies in a Semester-Long Course on the Use of GenAI. *J. Chem. Educ.* **2025**, *102*, 2507–2513.

22. Haraldsrud, A.; Odden, T. O. B. Bridging Artificial Intelligence and Real Intelligence: Self-Scaffolding Computational Modeling with Generative AI in Chemistry. *J. Chem. Educ.* **2025**, *102*, 4255–4266.

23. Cooper, M. M. A Preliminary Set of Principles to Support Learning in the Context of Generative AI. *J. Chem. Educ.* **2025**, ASAP Article. https://doi.org/10.1021/acs.jchemed.5c01413.

24. Clark, T. M.; Tafini, N. Design and Implementation of a General Chemistry Course that Promotes AI Use. *J. Chem. Educ.* **2025**, *102*, 4017–4023.

25. Krupp, L.; Steinert, S.; Kiefer-Emmanouilidis, M.; Avila, K. E.; Lukowicz, P.; Kuhn, J.; Küchemann, S.; Karolus, J. Unreflected Acceptance – Investigating the Negative Consequences of ChatGPT-Assisted Problem Solving in Physics Education. arXiv:2309.03087 [physics.ed-ph], 2023., https://arxiv.org/abs/2309.03087, (accessed Jan 10, 2025).

26. AI for Education. Prompt Framework for Educators: The Five "S" Model — AI for Education. https://www.aiforeducation.io/ai-resources/the-five-s-model, (accessed Jan 10, 2025).

27. Go Direct® SpectroVis® Plus Spectrophotometer - Vernier. https://www.vernier.com/product/go-direct-spectrovis-plus-spectrophotometer/, (accessed Jan 10, 2025).

28. Legacy Spectrometers — Ocean Optics. https://www.oceanoptics.com/resources/legacy-spectrometers/, (accessed Jan 10, 2025).

29. LabQuest 2 - Vernier. https://www.vernier.com/product/labquest-2/, (accessed Jan 10, 2025).

30. Stainless Steel Temperature Probe - Vernier. https://www.vernier.com/product/ stainless-steel-temperature-probe/, (accessed Jan 10, 2025).

31. Kennedy, A. B. W.; Sankey, H. R. The thermal efficiency of steam engines. Report of the committee appointed to the council upon the subject of the definition of a standard or standards of thermal efficiency for steam engines: with an introductory note. (Including appendixes and plate at back of volume). *Minutes of the Proceedings of the Institution of Civil Engineers* **1898**, *134*, 278–312.

32. Daniel, D.; West-Mitchell, K. The Sankey diagram: An exploratory application of a data visualization tool. *Transfusion* **2024**, *64*, 967–968.

33. Schmidt, M. The Sankey Diagram in Energy and Material Flow Management. *J. Ind. Ecol.* **2008**, *12*, 82–94.

34. Otto, E.; Culakova, E.; Meng, S.; Zhang, Z.; Xu, H.; Mohile, S.; Flannery, M. A. Overview of Sankey flow diagrams: Focusing on symptom trajectories in older adults with advanced cancer. *J. Geriatr. Oncol.* **2022**, *13*, 742–746.