Measuring Undergraduate STEM Reasoning: Graphical, Covariational and Proportional Skills as Predictors of Course Performance

Chukwuemeka Metu
The University of Alabama

Nigar Altindis
The University of Alabama

Kayode Obafemi
The University of Alabama

This study examined the role of students' graphical, covariational, and proportional (GCP) reasoning skills in predicting performance in introductory STEM courses. Specifically, it investigates whether GCP proficiency predicts overall course performance. Data from 382 students revealed significant course-level differences, $F(2, 379) = 10.27$, $p < .001$. Students in Introductory Chemistry for Engineers scored highest (EMM = 59.0), significantly outperforming peers in Biology (EMM = 49.5) and Chemistry (EMM = 43.9). GCP scores were moderately correlated with course performance ($r = .56$). Mixed-effects modeling confirmed that GCP scores were a strong predictor of final performance ($\beta = 0.025$, $p < .001$), even when accounting for variation across majors. This study has implications for STEM instructors. Assessments should be designed to evaluate students' application of knowledge, problem-solving skills, and ability to formulate coherent arguments. This would give a more comprehensive picture of a student's academic potential that goes beyond memory recall.

Keywords: Graphical Interpretations, Covariational Reasoning, and Proportional Reasoning.

Introduction

Science, Technology, Engineering, and Mathematics (STEM) courses are crucial for addressing the complex problems that plague today's society. They equip students with the knowledge and skills necessary to perform multiple tasks, collect and evaluate evidence, and make informed judgments about information (English, 2016; Topsakal et al., 2022; Whitacre, 2024). Research indicates that STEM graduates make significant contributions to increased economic growth and employment (Bacovic et al., 2022), underscoring their importance. Despite the benefits, success in introductory STEM courses remains a significant challenge, with many students struggling and ultimately leaving STEM majors (Rasmussen et al., 2019; Smith & Funk, 2021). A growing body of research suggests that students' difficulties may be linked to foundational reasoning skills necessary for engaging with STEM content. In particular, graphical interpretation, covariational reasoning, and proportional reasoning are essential for making sense of quantitative relationships, modeling dynamic systems, and applying mathematical concepts in scientific contexts (Altindis et al., 2024; Bowe et al., 2025; Rodriguez & Jones, 2024; Carlson et al., 2002, 2022; Moore, 2025). This study examines whether these forms of reasoning predict success in STEM courses.

Literature Review

Graphical interpretation, covariational reasoning, and proportional reasoning are foundational quantitative practices that underpin success in STEM learning. These forms of reasoning allow students to move beyond rote memorization by supporting their ability to analyze data, model relationships, and draw meaningful conclusions from quantitative information (Altindis, 2025; Emeji et al., 2025). For instance, graphical interpretation enables students to extract and synthesize information from visual representations, a skill necessary in disciplines ranging from biology to engineering (Bowe et al., 2025; Rodriguez & Jones, 2024). Covariational reasoning enables students to track how quantities change in relation to one another (Carlson et al., 2002).

Proportional reasoning enables the recognition and application of multiplicative relationships, which is essential for problem-solving across STEM contexts (Modestou & Gagatsis, 2009). Together, these reasoning practices provide students with the cognitive tools needed to interpret evidence, evaluate claims, and engage in scientific inquiry (Altindis et al., 2024; Altindis & Fonnger, 2025; Carlson et al., 2002, 2022; Rodrigues & Jones, 2024; Moore, 2025).

Graphical interpretation can be defined as a cognitive process through which students actively extract information, recognize patterns, and construct meaning from visual representations (Altindis et al., 2024; Bowe et al., 2025; Curcio, 1987; Glazer, 2011). The process of graph interpretation occurs through three steps: (1) recognizing visual patterns; (3) translating visual patterns into quantitative relationships; and (3) associating the relationships with the referents (Carpenter & Shah, 1998; Shah et al., 1999). While linear graphs are often manageable, interpreting them becomes more challenging as the graphs become more complex. As students struggle with translating nonlinear patterns into quantitative relationships and with identifying how the rate of change varies across different points of a function domain (Altindis, 2025; Altindis & Fonger, 2025; Altindis et al., 2024; Bowe et al., 2025; Carlson et al., 2002; Jones, 2019; Moore & Thompson, 2015; Moore, 2025; Roth & Bowen, 2001).

Covariational reasoning requires a high level of reasoning, as it can enhance students' comprehension of essential mathematical ideas (Thompson & Carlson, 2017). Covariational reasoning is a cognitive process that involves coordinating the values of two changing quantities and considering how they change in respect to one another (Altindis, 2025; Altindis & Fonger, 2025; Altindis et al., 2024; Bowe et al., 2025; Büchter et al., 2024; Carlson et al., 2002). A study by Carlson et al. (2002) revealed that with CR, students were able to create images of a function's dependent variable changing simultaneously with the imagined change of the independent variable, and in certain cases, images of the rate of change for continuous intervals of a function's domain. However, the students' abilities are sometimes misinterpreted by teachers who possess low-level covariational reasoning. When Zeytun et al. (2010) conducted a study on the mathematics teachers' reasoning levels, they found that some teachers have weak covariational reasoning, which leads to poor predictions about a student's ability. The degree to which covariational reasoning is fundamental in educational systems varies among countries (Thompson et al., 2017). The curriculum and Instruction in the United States, as indicated by Thompson and Carlson (2017), are not developing the covariational reasoning abilities of the students, thus resulting in many weaknesses in core mathematical concepts. Proportional reasoning is the ability to understand and work with multiplicative relationships between quantities (Modestou & Gagatsis, 2009). It involves knowing when two situations are proportional and using multiplication to solve problems instead of simple addition. Martínez Ortiz (2015) found that integrating LEGO robotics with mathematics instruction supported students' proportional reasoning by embedding ratio and proportion tasks within engineering design challenges.

Despite their importance, graphical, covariational, and proportional reasoning have often been studied in isolation rather than jointly as an interconnected set of quantitative practices. There is limited understanding of how these reasoning skills interact to shape students' STEM performance and whether they can serve as predictors of success across courses. This study addresses that gap by examining the role of these reasoning constructs in undergraduate STEM achievement, thereby contributing to both theory and practice in STEM education.

Aim of the Study

This study examines undergraduate students' graphical, covariational, and proportional (GCP) reasoning skill,—as measured by the GCP scale—in the context of introductory STEM courses. Specifically, it examines whether students' GCP reasoning proficiency predicts overall course performance. The two research questions guide this work: (1) Do students' GCP scores differ significantly across Introductory Biology, Introductory Chemistry, and Introductory Chemistry for Engineers? (2) To what extent do students' GCP skills predict their final performance in STEM courses?

Method

This study was conducted at a research-intensive university in the northeastern United States. The university is primarily comprised of White students (82%) and has a greater proportion of female students (56%) than male students (44%). Participants (n=382) were recruited from two sections of an introductory biology course required for life science majors, two sections of an introductory chemistry course required for life science and chemistry majors, and one section of an introductory chemistry course required for engineering majors. Data collected included pre- and post-assessments. There was no intervention for these courses; this is an exploratory study.

The GCP assessment scale consisted of 17 multiple-choice items: six testing covariational reasoning skills (Carlson et al., 2015), five focusing solely on graphical reasoning, and six assessing students' proportional reasoning skills. Each item was scored as correct (1) or incorrect (0). A percentage score was calculated for each student to represent overall proficiency in GCP reasoning skills. These students participated in three courses: Introductory Biology, Introductory Chemistry, and Introductory Chemistry for Engineers - taught by five instructors. Sample sizes varied across instructor–course combinations: Introductory Chemistry for Engineers with instructor A (n = 39/132, 30%), Introductory Chemistry with instructor B (n = 168/483, 35%), Introductory Chemistry with instructor C (n = 75/183, 41%), Introductory Biology with instructor D (n = 32/183, 18%), and Introductory Biology with instructor E (n = 68/181, 38%). These differences highlight the uneven distribution of students across courses and instructors. To address these disparities and variations in grading standards, students' final course grades were normalized. Specifically, scores were ranked within each course, converted to percentiles based on their relative standing, and then standardized using a normal distribution (mean = 0, SD = 1). This process ensured fair and consistent comparisons across courses and instructors, allowing student performance to be evaluated on a common scale.

Validity and Reliability

The content validity of the assessment was established through expert review by STEM education researchers in mathematics, biology, chemistry, and physics, who are subject matter specialists who confirmed the relevance, clarity, and representativeness of the items. In addition, several items were adapted and revised from previously validated instruments, including the *Calculus Concept Readiness (CCR)* assessment (Carlson et al., 2015), which was itself developed through research and expert input and has documented evidence of reliability and validity. Construct validity was further examined through Rasch analysis, with all 17 items demonstrating acceptable fit within the threshold of 0.6–1.4, thereby supporting the internal structure of the assessment. Item difficulty values ranged from –1.866 (easiest) to 2.026 (most difficult), reflecting a broad spectrum of student ability levels. The strong negative correlation ($r = –0.999$) between item difficulty and the proportion of correct responses confirmed the expected patterns. Reliability was evaluated using multiple indices of internal consistency, as shown in Table 1. All coefficients fell within the acceptable range of 0.70–0.80, indicating that the assessment demonstrates moderate and dependable reliability.

Table 1. Reliability Measures and Scores

Reliability Measures	Reliability Score
WLE (Weighted Likelihood Estimation) Reliability	0.736
EAP (Expected A Posteriori) Reliability	0.76
Cronbach Alpha	0.77

Results

Differences in GCP Scores Across Introductory STEM Courses

Table 2 below shows the differences in GCP scores across STEM courses. At the course level, students in Introductory Chemistry for Engineers achieved the highest GCP reasoning scores (Estimated Marginal means (EMM) = 59.0, SE = 3.23, 95% CI [52.6, 65.3]). Students in Introductory Biology scored moderately, with a combined mean of EMM = 49.5 across two instructors. In contrast, students in Introductory Chemistry had the lowest performance, with a combined mean of EMM = 43.9 across two instructors. These results indicate that students in Introductory Chemistry for Engineers outperformed their peers in both Introductory Biology and Introductory Chemistry on GCP measures.

Table 2. Estimated Marginal Means of GCP Scores by Course

Course	EMM	SE (range across instructors)	95% CI (range)
Introductory Chemistry for Engineers	59.0	3.23	[52.6, 65.3]
Introductory Biology	49.5	2.44–3.56	[42.9, 58.3]
Introductory Chemistry	43.9	1.56–2.31	[39.4, 47.8]

The analysis of variance (ANOVA) confirmed that the effect of course on GCP scores was statistically significant, $F(2, 379) = 10.27$, $p < .001$. ANOVA assumptions were assessed prior to analysis. Levene's test confirmed homogeneity of variance across groups, and the Q-Q plot indicated that residuals were approximately normally distributed, with only minor deviations at the tails. Thus, both assumptions of homogeneity and normality were satisfied. A Tukey post hoc test further confirmed that students in Introductory Chemistry for Engineers scored significantly higher than students in both Introductory Biology and Introductory Chemistry, whereas no significant difference was found between the latter two courses. GCP reasoning scores were moderately correlated with final course performance ($r = .56$), suggesting that stronger reasoning skills were consistently linked to higher achievement.

GCP Reasoning Skills as Predictors of Course Performance

We checked the normality assumption for the final course score and observed clustering in the data. Then, we examine a series of linear mixed-effects models. GCP pre-assessment scores were included as a fixed effect, while course, instructor, and major were initially treated as random effects. We examined the variability of random effects and conducted model comparison tests. As shown in Table 3, the standard deviations of the random effects for Instructor, Majors, and Course were generally small, with Course exhibiting the smallest values, suggesting a limited contribution to the overall variability. Bootstrap confidence intervals supported this finding, as the standard deviation for Course was nearly zero while Majors displayed slightly more variability. Given these patterns, we further compared models using both the Likelihood Ratio Test (LRT) and the Permutation-Based Test (PBtest). Results indicated that including Majors as a random effect significantly improved model fit (LRT = 4.893, p = 0.027; PBtest = 4.893, p = 0.019). Therefore, the final model retained Majors as a random effect, while Course and Instructor were omitted to achieve a better-fitting model.

Table 3. Random Effects and Model Selection Results

Effect	Std. Dev. (95% CI) – Normal Approximation	Std. Dev. (95% CI) – Bootstrap	Decision
Instructor	0.059 – 0.305	0.069 – 0.359	Omit
Majors	0.000 – 0.223	(–0.044) – 0.108 (p=.027 *)	Retain
Course	0.000 – 0.191	0.000 – 0.000	Omit
Residual	0.681 – 0.790	0.684 – 0.793	–
Intercept	–0.975 – –0.563	(–1.027) – (–0.579)	–
GCP Score	0.021 – 0.029	0.022 – 0.029	Fixed effect retained

**Note: LRT = Likelihood Ratio Test; PBtest gave similar results (stat = 4.893, p = 0.019)*

We examine the extent to which students' GCP skills predicted final performance in STEM courses, while also accounting for variation across majors. As shown in Table 4, the GCP score was a strong and statistically significant predictor of students' normalized final scores. This indicates that for each one-unit increase in GCP scores, students' normalized final scores increased by approximately 0.025 units, holding other factors constant. Thus, GCP skills demonstrated a consistent relationship with STEM course performance.

Table 4. Final Mixed-Effects Model Estimates Predicting Normalized Final Scores

Parameter	Estimate	Std. Error	95% CI (Lower, Upper)	Variance	Std. Dev.	df	t	p
Fixed Effects								
Intercept	–0.769	0.105	[–0.98, –0.57]	–	–	161.91	–7.29	< .001
GCP Score	0.0248	0.0019	[0.021, 0.029]	–	–	367.09	12.94	< .001
Random Effects								
Majors (Intercept)	–	–	[0.07, 0.37]	0.03	0.18	–	–	–
Residual	–	–	[0.68, 0.79]	0.54	0.73	–	–	–

Discussion

This study revealed notable differences in GCP reasoning among introductory STEM courses. Students who enrolled in Introductory Chemistry for Engineers scored significantly higher than students in both Introductory Biology and Introductory Chemistry. These findings suggest that engineering-oriented chemistry courses may place greater emphasis on reasoning practices or attract students with stronger quantitative preparation. Lower performance in

Introductory Chemistry and Biology courses may point to missed opportunities to engage students with reasoning practices as central tools for inquiry. Prior research has indicated that students often struggle with graph interpretation and rate-of-change reasoning when these skills are not explicitly scaffolded (Moore & Thompson, 2015; Roth & Bowen, 2001). Thus, the results highlight the importance of intentional instructional design that develops quantitative reasoning across STEM disciplines, not only in engineering contexts.

We found that GCP skills were a strong predictor of final course performance across all introductory STEM contexts. The final mixed-effects model indicated that for each unit increase in GCP scores, students' normalized final scores increased by approximately 0.025 units, even after accounting for variation across majors. This finding highlights the crucial role of reasoning skills in fostering academic success in STEM. It aligns with earlier work demonstrating that covariational reasoning is critical for understanding functions, modeling phenomena, and solving quantitative problems (Altindis & Fonger, 2025; Altindis, 2025; Altindis et al., 2024; Bowe et al., 2025; Carlson et al., 2002). The moderate correlation between GCP and course performance ($r = .56$) further reinforces this relationship, suggesting that reasoning proficiency can serve as an early diagnostic indicator of students' potential achievement.

The model also revealed that major-level differences contributed modest variability, while course- and instructor-level effects were negligible. This suggests that students' disciplinary orientation and preparation may matter more for performance than differences in course delivery. However, further research is needed to understand why these differences emerge, how majors influence students' reasoning development, and whether specific curricular pathways or instructional approaches account for the observed variability.

Implications

The results carry several implications for STEM education. First, they emphasize the need for instructional strategies that explicitly develop graphical, covariational, and proportional reasoning across disciplines. Biology and chemistry courses, in particular, may benefit from integrating reasoning tasks into laboratory and lecture settings to promote deeper conceptual engagement. Second, the strong predictive role of GCP skills suggests that reasoning-based assessments could serve as early indicators of student achievement, providing opportunities for targeted support. Finally, embedding reasoning tasks in authentic, applied contexts that are similar to those found in engineering courses may help foster quantitative practices across STEM.

Limitations and Future Directions

While this study provides evidence for the importance of GCP reasoning, several limitations should be noted. First, the data were drawn from a single institutional context, which may limit generalizability. Second, although the study controlled for majors, other factors, such as prior mathematical preparation and demographic variables, were not included. Future research should explore these additional influences, as well as examine how instructional practices within biology and chemistry could be redesigned to support reasoning more effectively. Ultimately, longitudinal research is necessary to investigate how GCP skills evolve and whether targeted interventions can bridge performance gaps across courses.

Acknowledgments

This research was supported by the National Science Foundation, Award No. 2013427. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Altindis, N., & Fonger, N. (2025). A culturally inclusive mathematics learning environment framework: Supporting students' representational fluency and covariational reasoning. *Education Sciences, 15*(8). https://doi.org/10.3390/educsci15080980

Altindis, N., Bowe, K. A., Couch, B., Bauer, C. F., & Aikens, M. L. (2024). Exploring the role of disciplinary knowledge in students' covariational reasoning during graphical interpretation. *International Journal of STEM Education, 11*(1). https://doi.org/10.1186/s40594-024-00492-5

Altindis, N. (2025). Networking theories of quantitative reasoning and mathematical reasoning to explore students' understanding of functions. *The Journal of Mathematical Behavior, 80,* 101276. https://doi.org/10.1016/j.jmathb.2025.101276

Bacovic, M., Andrijasevic, Z., & Pejovic, B. (2022). STEM education and growth in Europe. *Journal of the Knowledge Economy, 13*(3), 2348-2371. https://doi.org/10.1007/s13132-021-00817-7

Bowe, K. A., Altindis, N., Bauer, C. F., & Aikens, M. L. (2025). *Cognitive resources drawn from different STEM disciplines for interpreting a titration experiment logistic graph: A case study of student thinking*. *Disciplinary and Interdisciplinary Science Education Research, 7*(20). https://doi.org/10.1186/s43031-025-00140-0

Büchter, T., Eichler, A., Böcherer-Linder, K., Vogel, M., Binder, K., Krauss, S., & Steib, N. (2024). Covariational reasoning in Bayesian situations. *Educational Studies in Mathematics*, *115*(3), 481-505.

Carlson, M. P., Madison, B., & West, R. D. (2015). A study of students' readiness to learn calculus. *International Journal of Research in Undergraduate Mathematics Education*, *1*(2), 209-233. https://doi.org/10.1007/s40753-015-0013-y

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education*, *33*(5), 352-378. https://doi.org/10.2307/4149958

Carlson, M.P., O'Bryan, A., Rocha, A. (2022). Instructional conventions for conceptualizing, graphing and symbolizing quantitative relationships. In: Karagöz Akar, G., Zembat, İ.Ö., Arslan, S., Thompson, P.W. (eds) Quantitative Reasoning in Mathematics and Science Education. Mathematics Education in the Digital Era, vol 21. Springer, Cham. https://doi.org/10.1007/978-3-031-14553-7_9

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75. https://doi.org/10.1037/1076-898X.4.2.75

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393. https://doi.org/10.5951/jresematheduc.18.5.0382

Emeji, E. I., Nwoye, M. N., Aniaku, O. L., Okeke, A. M., Ngwu, A. N., Osakwe, I. J., Inweregbuh, O. C., Onyeagba, M. C. N., Nwikpo, M. N., Eluemuno, A. I., Ezeonwumelu, V.

U., Metu, C. A., & Anyanwu, A. N. (2025). Influence of students' reasoning patterns on achievement in senior secondary school plane geometry: Implications for biology education practitioners. *Cineforum, 65*(3), 147–162.
English, L. D. (2016). STEM education K-12: perspectives on integration. *International Journal of STEM Education, 3*(3). https://doi.org/10.1186/s40594-016-0036-1
Glazer, N. (2011). Challenges with graph interpretation: a review of the literature. *Studies in Science Education*, *47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307
Jones, S. R. (2019). Students' application of concavity and inflection points to real-world contexts. *International Journal of Science and Mathematics Education, 17*, 523-544. https://doi.org/10.1007/s10763-017-9876-5
Martinez Ortiz, A. (2015). Examining students' proportional reasoning strategy levels as evidence of the impact of an integrated LEGO robotics and mathematics learning experience. *Journal of Technology Education, 26*(2), 46–69.
Modestou, M., & Gagatsis, A. (2009). Proportional reasoning: The strategies behind the percentages. *Acta Didactica Universitatis Comenianae. Mathematics, 9*, 25–40.
Moore, K. (2025). A Framework for Time and Covariational Reasoning. *The Mathematics Educator*, *33*(1).
Moore, K. C., & Thompson, P. W. (2015, February). Shape thinking and students' graphing activity. In *Proceedings of the 18th meeting of the MAA special interest group on research in undergraduate mathematics education* (pp. 782-789). Pittsburgh: RUME.
Rodriguez, J. M. G., & Jones, S. R. (2024). How students understand graphical patterns: Fine-grained, intuitive knowledge used in graphical thinking. *Journal for Research in Mathematics Education*, *55*(2), 96-118. https://doi.org/10.5951/jresematheduc-2022-0167
Roth, W. M., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education, 32*(2), 159-194. https://doi.org/10.2307/749672
Şen Zeytun, A., Çetinkaya, B., & Erbaş, A. K. (2010). Mathematics teachers' covariational reasoning levels and predictions about students' covariational reasoning abilities. *Educational Sciences in Theory and Practice, 10*(3), 1601-1612. https://hdl.handle.net/11511/97522
Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91*(4), 690–702. https://psycnet.apa.org/doi/10.1037/0022-0663.91.4.690
Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.
Thompson, P. W., Hatfield, N. J., Yoon, H., Joshua, S., & Byerley, C. (2017). Covariational reasoning among US and South Korean secondary mathematics teachers. T*he Journal of Mathematical Behavior, 48*, 95–111. https://doi.org/10.1016/j.jmathb.2017.08.001
Topsakal, İ., Yalçın, S. A., & Çakır, Z. (2022). The effect of problem-based STEM education on the students' critical thinking tendencies and their perceptions for problem solving skills. *Science Education International*, *33*(2), 136-145. https://doi.org/10.33828/sei.v33.i2.1
Whitacre, P. (2024). Supporting K-12 STEM Education to Create the Foundations for Innovation: Proceedings of a Workshop--In Brief. *National Academies Press*.