Promoting Multivariable Causality Reasoning Through Simulations: The Role of Exploratory Learning in Youth-at-risk Schools

Janan Saba [0000-0001-9211-3114], Malak Abu Rmaileh, Tali Gal [0000-0002-621-4672]

Hebrew University of Jerusalem, Jerusalem, Israel
Janan.saba@mail.huji.ac.il, malak.aburmaileh@mail.huji.ac.il tali.gal@mail.huji.ac.il

Abstract. While it is common in science education to use interactive simulations to support exploratory learning of complex phenomena, students struggle to make sense of and reason about these complex phenomena. This challenge partly stems from the need to understand how multiple interacting factors produce observed outcomes, a process referred to as Multivariable Causality reasoning (MVC reasoning). In this study, we examine the applicability of the Problem-Solving before Instruction (PS-I) approach in this context. In PS-I, learners are given complex tasks that aid their comprehension of the domain before receiving instruction on the target concepts. This study explored the development of MVC reasoning through interactive simulations utilizing the PS-I approach with students from youth-at-risk schools. At-risk students, defined as those who have dropped out of traditional schools, are likely to fail in school and not complete certain levels of education. High-school students (N=197) from both mainstream and youth-at-risk schools were randomly assigned to either an exploration-first condition or an instruction-first condition to learn about Predator-Prey relationships. They completed a pretest, followed by an intervention comprising an exploration task (Task 1) using a simulation before instruction (exploration-first) or after (instruction-first condition) the MVC concepts were taught. This was succeeded by an exploration task (Task 2) on the same topic and a posttest. Findings reveal a compelling and unexpected outcome: engagement with interactive simulations seems to yield greater benefits for at-risk students compared to their peers in mainstream schools, regardless of the learning approach adopted. For future work, the study suggests the integration of an AI-data-driven scaffold, in terms of an intelligent system that is capable of analysing students' real-time interactions within the simulation, identifying instances where students struggle in managing the multiple variables, and providing tailored strategic guidance aimed at facilitating deeper engagement in MVC.

Keywords: Multivariable causality reasoning, Exploratory learning, Interactive simulation, At-risk students.

1 Introduction

Inquiry-based learning is a fundamental approach in science education, designed to enhance students' scientific reasoning skills [1,2]. This approach is particularly effective in exploring authentic scientific phenomena, where multiple variables interact and influence the outcomes. The strategy of investigating complex relationships among factors within these scientific phenomena is called the Multivariable Causality (MVC) [3]. MVC reasoning is a scientific skill that addresses why and how various factors affect outcomes. It refers to the cognitive ability of learners to comprehend, analyze, and reason about complex phenomena where multiple factors act and interact to give rise to an observed outcome. For example, when investigating the spread of a pandemic, the impact of the vaccination rate on pandemic spread depends on other factors such as the transmission rate of the virus in the population, individuals' recovery rate, etc. Studies have found it challenging for students to describe and reason about the nonlinear relationship between variables in complex systems [4].

To help learners develop a comprehensive understanding of MVC, we are building on previous studies that suggested using computer-based interactive simulations [5,6]. Interactive simulations are visual representations of complex phenomena that encompass events and processes [7], offering a means to explore the various relationships inherent within complex phenomena [5, 6, 8, 9] by systematically manipulating multiple variables in the simulation and observing the outcomes through visual graphical representations [7, 8]. Studies have shown that simulation-based learning significantly enhances students' conceptual understanding and learning transfer [5, 6]. However, few studies have examined the development of MVC reasoning using interactive simulations. Some of these studies were conducted with middle- to high-school students in mainstream educational settings [3] and with undergraduate students [10]. Thus, the first gap identified is that, to the best of our knowledge, no study has explored the development of MVC reasoning through interactive simulations among underrepresented students, particularly those from youth-at-risk schools, schools that serves students who are at risk of not meeting academic or social expectations, or who may struggle to graduate high school. Thus, At-risk students are those who have dropped out of mainstream schools and are likely to fail in school and not complete a certain level of education [11,12]. They need support in developing effective learning strategies, and technology can significantly help them in this endeavor [13]. Research indicates that technology not only enhances students' self-efficacy, self-confidence, and motivation as they successfully complete designated tasks [14] but also has the potential to improve their overall learning outcomes [15]. Despite the potential benefits of technology-based learning for at-risk students, evidence of these benefits is still lacking, particularly regarding the integration of interactive simulations. Additionally, the second identified gap is that it remains unclear how the simulations should be used alongside instruction to better support the learning of at-risk students.

To address these two gaps, we investigated the applicability of exploratory learning by engaging at-risk students in inquiry learning through interactive simulations while learning science. For this aim, we are based on Problem-Solving before Instruction approach [16], where learners are given complex tasks that help them make sense of the domain before receiving instruction on the target concepts. In the context of this study, we explored the effectiveness of the exploration-first approach in terms of engaging in an exploration task using simulations prior to receiving direct instruction on MVC reasoning. This approach was compared with instruction-first approach, which starts with direct instruction of MVC, followed by an exploration task using simulation. It is based on Instruction prior to Problem-Solving (I-PS), where students receive direct instruction on the target concepts, followed by implementation through Problem-Solving. In this study, the two approaches were compared in two educational settings (mainstream and youth-at-risk).

2 Goals and research questions

The purpose of this study is to investigate the impact of adopting the exploration-first approach using interactive simulations on enhancing learning about MVC among students in youth-at-risk high schools. Thus, the effectiveness of two learning approaches, exploration-first versus instruction-first, was compared across two types of schools: youth-at-risk schools and mainstream schools. This study poses the following question:

RQ. What is the impact of engaging in exploratory learning through interactive simulations on the development of MVC reasoning among students from youth-at-risk schools compared to those from mainstream schools, across two learning approaches, exploration-first and instruction-first?

3 Method

3.1 Participants

A total of 197 male high school students (aged 16-18) from the 9th, 10th, 11th, and 12th grades participated in the study at two urban schools with middle to low socioeconomic status. Students from the youth-at-risk school were randomly assigned to either the exploration-first condition (n = 50) or the instruction-first condition (n = 45). Similarly, students from the mainstream school were randomly assigned to either the exploration-first condition (n = 60) or the instruction-first condition (n = 42). All students and their parents provided consent to participate in the study according to the ethical requirements of the Ministry of Education. It's important to note that we included students from all high school grades, since MVC skills aren't typically part of the high school science curriculum. Thus, we treated grade level as a fixed independent variable in all statistical analyses to account for any confounding effects on learning outcomes.

3.2 Design and procedure

The study employed a controlled-experimental design with a pretest-intervention-posttest procedure. Students completed a predation pretest, followed by an intervention that involved an exploration task (Task 1) regarding predation using the "Wolf Sheep Predation" simulation and instruction on MVC. After the intervention, all groups engaged in an exploration task (Task 2) using the "Wolf Sheep Predation" simulation and completed a posttest on Task 2. Table 1 illustrates the design of the study for the four groups. The total duration of the study was 1.5 hours.

Table 1. The sequence of learning design for each group

Education setting	Learning approach	Pretest	Intervention		Posttest
Youth-at-risk	Exploration-first	Pretest	Task1	Instruction	Task 2+ posttest
	Instruction-first		Instruction	Task 1	
Mainstream	Exploration-first		Task1	Instruction	
	Instruction-first		Instruction	Task 1	

3.3 Materials

Learning materials included two exploration tasks and instruction: Task 1 and Task 2 targeted the exploration of a problem related to wolf and sheep predation within an ecosystem using the “Wolf Sheep Predation” simulation. It is an online interactive simulation that was developed by combining the “Wolf Sheep Predation” NetLogo model [17] and the NetTango platform [18]. Students were asked to explore the relationships and interactions among three factors inside an ecosystem, which includes two populations, wolves and sheep (sheep reproduce, wolves reproduce, and wolves gain energy from food; Figure 1). The simulation allows students to conduct several experiments iteratively and investigate the relationships between the three variables and how they jointly affect the outcome. Each task includes an explanation question that students were asked to provide their reasoning related to whether and how different factors could affect the stability of the system (Task 1) and the maximum number of wolves in a stable system (Task 2). The instruction phase illustrated phenomena consisting of multivariable interactions. It focused on exploring and reasoning about how nonlinear relationships between several variables give rise to the observed outcomes.

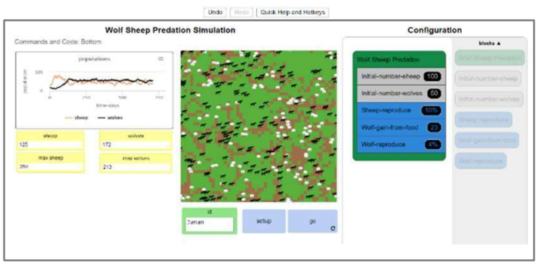

Figure 1. Wolf Sheep Predation simulation.

3.4 Data Sources and Analysis

Two data collection tools were used in the study: (1) Pretest consisted of one open-ended item that asked students to describe factors that may affect the ecosystem of sheep and wolves, and to explain how these factors may affect the stability of the ecosystem. It aims to test students' prior knowledge of MVC reasoning before the intervention. (2) Posttest consisted of one open-ended item that asked students to provide reasoning related to whether and how different factors could affect the maximum number of wolves in a stable ecosystem based on their exploration of the simulation in Task 2.

To investigate students' MVC reasoning, their responses to the open-ended items on the pretest and posttest were analyzed using a coding scheme developed based on Krist et al. [19] and Saba et al. [10], along with a bottom-up analysis of students' responses. The coding scheme comprised three codes and their combinations: (1) Describing: a description of the relationship between factors at the population level and their effect on an outcome at this level, without explaining how these factors act and interact. (2) Unpacking: describing and explaining how factors act and interact at the population level and how these factors affect outcomes at the population level. (3) Linking: connecting the unpacked or described factors at the population level to an outcome at the observed level. A coding scheme for students' responses was created based on the components and the number of factors used; for example, D1 – describing one factor, D1L: Describing and Linking one factor. Scores range from 0 to 12: None – 0; D1 – 1; D2 – 2; D3 – 3; D1L – 4; D2L – 5; D3L – 6; U1 – 7; U2 – 8; U3 – 9; U1L – 10; U2L – 11; U3L – 12. Following, students were provided with scores in the pretest and posttest.

4 Results

Pre-Intervention difference. Levene's Test was first conducted to assess the assumption of homogeneity of variance. Given that Levene's Test was significant, $F(3, 193) = 11.88$, $p < .001$, indicating unequal variances across groups, Welch's ANOVA was employed as a robust alternative to the traditional three-way ANOVA. Results of the Welch's ANOVA revealed a significant difference between school types, $F(1, 196) = 5.80$, $p = .017$, indicating that students in youth-at-risk school and mainstream school differed significantly at pretest. However, no significant differences were observed for instructional approach, $F(1, 196) = 0.02$, $p = .903$, or grade level, $F(1, 196) = 1.63$, $p = .188$. Table 2 presents the means and standard deviations of the pretest across schools and learning approaches.

Learning gain from pre- to posttest. To examine the effect of school type and learning approach on students' improvement from pre- to posttest across school and learning approach, A 2 (Time: pretest, posttest) x 2(School) x 2(Learning approach) repeated measures ANCOVA statistical test was conducted to examine students' learning gain related to MVC reasoning from pretest to posttest. Time was treated as a within-subjects factor, while school and learning approach were between-subjects factors. The grade level was included as a covariate. Findings revealed a significant main effect only for Time with a small effect size, indicating that students improved their MVC reasoning from pretest to posttest, $F(1, 192) = 8.949$, $p = 0.003$, $\eta_p^2 = 0.045$. However, Findings regarding interactions effects revealed a significant interaction effect only for Time x School with a medium effect size, $F(1, 192) = 11.122$, $p = 0.001$, $\eta_p^2 = 0.055$ indicating that students from youth-at-risk demonstrated significantly superior improvement in their MVC reasoning from pretest to posttest compared to students from the mainstream school regardless learning approach and grade level (Figure 2).

Table 2. Means and Standard Deviations for pretest and posttest by school type and learning approach

School Type	Learning approach	N	Pretest M (SD)	Posttest M (SD)
At-risk	Exploration-first	50	1.28(1.67)	4.12 (3.68)
At-risk	Instruction-first	45	0.87(0.97)	4.62 (3.34)
Mainstream	Exploration-first	60	1.60(2.68)	2.98 (3.40)
Mainstream	Instruction-first	42	2.17(2.95)	3.62 (3.20)

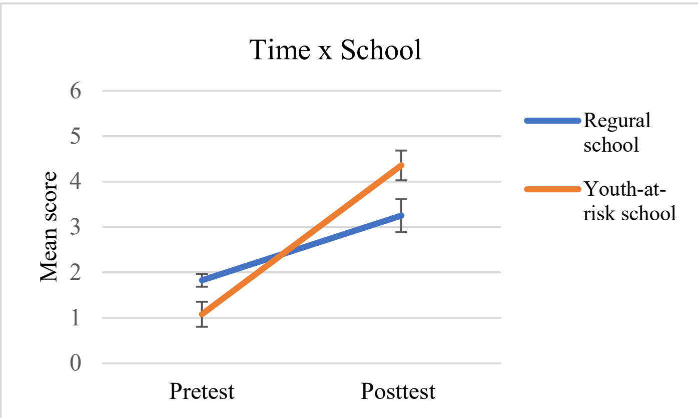

Figure 2. The interaction between Time and School from pretest to posttest

5 Scholarly Significant

This study contributes to the existing literature on the discussion of cultivating general domain science skills among youth-at-risk schools through exploratory learning. It addresses critical theoretical gaps concerning the effectiveness of technology-based learning interventions for at-risk students, particularly in engaging them through exploratory learning that utilizes interactive simulations. This study is the first to demonstrate that at-risk students can significantly benefit from an approach that combines exploratory learning with interactive simulations to enhance their scientific skills, despite having significantly lower prior knowledge of MVC compared to students in the mainstream school.

Notably, our findings reveal a compelling and unexpected outcome: engagement with interactive simulations seems to yield greater benefits for at-risk students compared to their peers in mainstream schools. These results may be more closely linked to prolonged socio-educational neglect and systemic inequities rather than low cognitive abilities. They can be attributed to the necessity for at-risk students to participate in a more stimulating and motivational learning environment that encourages self-expression and fosters a sense of agency in their educational experiences. Moreover, this underscores the importance of providing appropriate pedagogical support and enriching learning environments, which can enable at-risk students to realize their potential for significant academic achievement. In light of this context, this study suggests the integration of an AI-data-driven scaffold. This involves the development of an intelligent system that is capable of analyzing students' real-time interactions within the simulation environment. The system is designed to identify instances where students struggle in managing the multiple variables involved in the system. Moreover, it provides tailored strategic guidance aimed at facilitating deeper engagement in MVC.

Regarding the most effective learning approach for integrating interactive simulation, our results indicate that the sequence of the learning approach, whether exploration occurs before instruction or instruction is followed by exploration, does not significantly impact learning outcomes for students in either group.

While these results are preliminary, they pave the way for further discussion on the effects on students' performance in a transfer task, where they were introduced to a new task in a different context and used a different simulation. Additionally, we will examine the intermediate knowledge [20] gained in Task 1 during the intervention and its impact on students' performance in the posttest, exploring how this knowledge may serve as a predictor for subsequent learning outcomes. Practically, this study will provide significant insights for teachers, suggesting the integration of interactive simulations as a core element in designing scientific learning environments aimed at supporting at-risk students.

6 References

1. Wu, H., Wu, P., Zhang, W., & Hsu, Y. (2013). Investigating college and graduate students' multivariable reasoning in computational modeling. Science Education, 97(3), 337– 366. https://doi.org/10.1002/sce.21056
2. Stender, A., Schwichow, M., Zimmerman, C., & Härtig, H. (2018). Making inquiry-based science learning visible: the influence of CVS and cognitive skills on content knowledge learning in guided inquiry. International Journal of Science Education, 40(15), 1812-1831.
3. Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariable thinkers. Cognitive Development, 35, 92-110.

4. Liang, Y., Gao, Z., Gao, J., Wang, R., Liu, Q., & Cheng, Y. (2020). A new method for multivariable nonlinear coupling relations analysis in complex electromechanical system. Applied Soft Computing, 94, 106457. https://doi.org/10.1016/j.asoc.2020.106457
5. Roll, I., Butler, D., Yee, N., Welsh, A., Perez, S., Briseno, A., & Bonn, D. (2018). Understanding the impact of guiding inquiry: The relationship between directive support, student attributes, and transfer of knowledge, attitudes, and behaviors in inquiry learning. Instructional Science, 46(1), 77–104.
6. Saba, J., Hel-Or, H., & Levy, S. T. (2023). Much. Matter. in. Motion: learning by modeling systems in chemistry and physics with a universal programming platform. Interactive Learning Environments, 31(5), 3128-3147.
7. Moser, S., Zumbach, J., & Deibl, I. (2017). The effect of metacognitive training and prompting on learning success in simulation-based physics learning. Science Education, 101(6), 944-967.
8. de Jong, T. (1991). Learning and instruction with computer simulations. Education & Computing, 6, 217–229.
9. Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. Science, 322(5902), 682–683.
10. Saba, J., Kapur, M., & Roll, I. (2025). Learning about multivariable causality with interactive simulations: exploration before instruction may hurt immediate gains but benefits transfer. *Instructional Science*, 1-30.
11. Kaufman, P., Bradbury, D., Riley, R. W., Takai, R., & Owings, J. (1998). Characteristics of At-Risk Students in NELS:88. NATIONAL CENTER FOR EDUCATION STATISTICS. https://nces.ed.gov/pubs92/92042.pdf
12. McMillan, J. H., & Reed, D. F. (1994). At-Risk students and resiliency: factors contributing to academic success. The Clearing House a Journal of Educational Strategies Issues and Ideas, 67(3), 137–140. https://doi.org/10.1080/00098655.1994.9956043
13. Hannafin, K. M. (1991). TECHNOLOGY AND THE SUPPORT OF AT-RISK STUDENTS. The Journal of General Education, 40, 163–179. http://www.jstor.org/stable/27797135
14. Li, X., Zhang, J., & Yang, J. (2024). The effect of computer self-efficacy on the behavioral intention to use translation technologies among college students: Mediating role of learning motivation and cognitive engagement. Acta Psychologica, 246, 104259.
15. Bakar, K. A., Ayub, A. F. M., Luan, W. S., & Tarmizi, R. A. (2010). Exploring secondary school students' motivation using technologies in teaching and learning mathematics. Procedia - Social and Behavioral Sciences, 2(2), 4650–4654. https://doi.org/10.1016/j.sbspro.2010.03.744
16. Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. Educational Psychology Review, 29(4), 693– 715. https://doi.org/10.1007/s10648-016-9379-x.
17. Wilensky, U. (1999). NetLogo models library [Computer software]. In Center for connected learning and computer-based modeling. Northwestern University. Available from http://cclnorthwestern.edu/netlogo/models/.
18. Horn, M., Baker, J. & Wilensky, U. (2020). NetTango Web 1.0alpha. [Computer Software]. Evanston, IL. Center for Connected Learning and Computer Based Modeling, Northwestern University. http://ccl.northwestern.edu/nettangoweb/.
19. Krist, C., Schwarz, C. V., & Reiser, B. J. (2018). Identifying essential epistemic heuristics for guiding mechanistic reasoning in science learning. Journal of the Learning Sciences, 28(2), 160–205. https://doi.org/10.1080/10508406.2018.1510404

20. Loibl, K., Leuders, T., Glogger-Frey, I., & Rummel, N. (2024). CID: a framework for the cognitive analysis of composite instructional designs. Instructional Science, 1-25.