“A Fresh Squeeze on Data”: Exploring Gender Differences in Self-Efficacy and Career Interest in Computing Science and Artificial Intelligence among Elementary Students

Shuhan Li[1[0000-0001-9174-3691]], Annabel Hasty[2[0009-0002-0753-7508]], Eryka Wilson[3[0000-0002-2023-3608]], Roozbeh Aliabadi[4]

[1] Columbia University, New York, NY 10025
[2] Quest Academy, Palatine, IL 60067
[3] Keiser University, FL 33309
[4] ReadyAI, Pittsburgh, PA 15206

Abstract. Artificial intelligence (AI) is a growing field in both global job markets and educational spaces. This non-experimental quantitative study aims to explore how the educational program *A Fresh Squeeze on Data* affects students’ self-efficacy and career choices and whether gender will differentiate the learning outcomes. Under the social cognitive career theory framework, this study designs questionnaires as data collection instrument. The results suggest that the program significantly improves students’ comfortability with AI-related subjects but not for career interest or other measurements in self-efficacy. Unexpectedly, the program’s effect is not divided by gender. Nevertheless, this study opens up conversations about assisting students from underrepresented backgrounds to envision success in AI courses and career pathways through an activity-driven curriculum. The paper also informs educators and researchers to devise culturally responsive pedagogy in teaching AI that empowers young girls before they develop a gendered career view.

Keywords: K-12 Education, AI Education, Gender Equity, Self-Efficacy, Career Interest.

1 Introduction

The increasing importance of AI in major industries calls for relevant education among young learners. To prepare for their digitized future, students must cultivate data literacy earlier in their education. Educational institutes, in response, are gradually integrating relevant curricula into the classroom so that they can build up students’ computing skills and familiarize them with emerging AI technologies.

However, gender inequality in AI-related fields still lingers. Women’s underrepresentation can be explained by the dearth of female students in advanced computing science or AI-related courses. Despite many female students’ exposure to data-driven coursework in middle or high school, a majority discontinue their pursuit in these fields later in their life [4].

This non-experimental quantitative study delves into the effectiveness of a culturally responsive curriculum. From a socio-cognitive and pedagogical standpoint, this study will advocate for the active participation in school-led learning programs and the awareness of gender inclusivity in the design of AI educational materials.

2 Research Purposes and Questions

It is against the background of AI's rising importance and women's underrepresentation in the industry that this study is initiated. This paper explores the meaningfulness of a culturally responsive AI curriculum. It informs teachers of an innovative pedagogy that can both enhance AI content knowledge and achieve girls' empowerment. This paper asks the following research questions to navigate the subsequent analysis:

- RQ1: Does the participation in AI education programs impact students' self-efficacy and career interest in related fields?
- RQ2: To what extent are self-efficacy and career interest changed?
- RQ3: Does gender differentiate the changes in self-efficacy and career interest?

3 Social Cognitive Career Theory (SCCT)

SCCT provides the theoretical framework for this research. It assumes that individuals, their behaviors, and contexts jointly and bi-directionally affect each other [1, 8]. Self-efficacy is an important aspect in the SCCT circuit. It measures an individual's perceived competency in goals achievements [7]. This research also adopts the more unified model that takes demographics such as gender into consideration [11]. Social norms are part and parcel to shaping one's efficacy level. Hypothetically, women's low participation rate in AI-related fields can be attributed to inadequate learning opportunities and non-inclusive learning environment. These factors lower female learners' self-efficacy, thus preventing them from further career pursuits [5, 6].

4 Program Description

This study involves *A Fresh Squeeze on Data*, a student-centered, collaborative intervention program about data collection, data bias, and AI. Named after a picture book about problem-solving with data, this educational program encourages young students to be gamechangers and problem-solvers using the emerging technology of AI [10]. The children's book introduces a young girl who learns data collection from her mother, a data scientist, and reduces biases to solve problems for her community.

A Fresh Squeeze on Data lesson plan combines concepts with practices, leading students to effectively translate what they had learned to the real world. Each lesson lasts 90 minutes, consisting of two major hands-on activities. Through real-world simulations like running a candy stand and selecting an adopted pet, students are introduced to data collection, analysis, and bias identification [11]. These activities are

combined with small group discussions to facilitate peer cooperations. The learning materials are accessible to Grade 3 to 5 students.

The program adopts a constructivist pedagogy. A wide range of activities hone students' teamwork, problem-solving, and inquiry skillsets [3]. Students directly interact with basic AI products such as Teachable Machine to train an AI model for categorization. This exercise familiarizes students with machine learning. Simple as the activity seems, it breaks down the dense technical knowledge into straightforward examples. In this way, students explore AI with greater ease and thus develop more interest in the subject.

Contrary to a conventional program with only a final project, this course places multiple hands-on activities in varying difficulties throughout the entire learning period [3]. Accordingly, students dedicate more time engaging with the learning materials. Each activity involves open-ended discussion questions [11]. This enables students to explore the technology through inquisitive trial and error. Compared to the lecture-style classrooms, *A Fresh Squeeze on Data* encourages students to demonstrate their agencies to learn and think critically about AI applications.

5 Method

5.1 Survey Design

Inspired by the survey design of previous literature [2, 9], this study utilizes similar instruments. Students are given the same questionnaire before and after the training program. The difference in their answers reflects the impact of the program on student outcomes. The survey contains five constructs, three of which measure self-efficacy and two measure career interest. Each construct contains two prompts that ask students to rate their opinion on a scale of 1 (strongly disagree) to 5 (strongly agree). The self-efficacy constructs gauge students' interest in, comfort with, and attitudes towards learning about AI and computing science. The career interest constructs investigate students' knowledge of and propensity towards occupations related to AI and data science. The last section asks for students' gender identification, age, and previous AI and data science experience.

5.2 Sample and Hypotheses

This study involves a sample size of 48 students (n=48) who took part in the training program over the summer. Among them, 28 are male and 20 are female. Their ages range from 7 to 11. They come from diverse ethnic backgrounds. All sampled students have a certain degree of previous exposure in AI-related areas in either on-campus or extracurricular contexts. On average, the sampled students take AI- or STEM-related classes twice a week.

A repeated measures MANOVA is run on SPSS software to generate descriptive data on the survey results. Gender is the between-subject factor, while time is the within-subject factor. The dependent variables are the overall self-efficacy and career

interest level for each student. Cumulative self-efficacy is calculated by adding the scores of the first five prompts, subtracting that of the sixth. Cumulative career interest is the total score of the last four prompts. The following hypotheses are tested at a confidence level of 0.95:

- H1: *A Fresh Squeeze on Data* impacts students' self-efficacy level and career interest in AI or computing science.
- H2: Gender differentiates students' change in self-efficacy level and career interest in AI or computing science.

6 Data Analysis

6.1 Impact on Self-Efficacy

The mean for the overall self-efficacy level increases from 15.96 (sd = 4.156) to 16.48 (sd = 4.443) after the program. However, this increase is not statistically significant ($p = 0.090 > 0.050$). The overall effect of the program on students' self-efficacy level remains unclear.

Nonetheless, the results yield a significant change in students' answers to Prompt 4 (comfortability with learning AI and computing science), with a p-value of $0.010 < 0.050$. The overall score increases from 3.52 (sd = 1.091) to 3.85 (sd = 1.148). This indicates that *A Fresh Squeeze on Data* reduces students' stress about learning AI-related knowledge. Part of the reason is that the course is conducted in an interactive environment using a child-friendly story book. The increased comfortability, in return, helps boost students' confidence in their learning ability.

6.2 Impact on Career Interest

The mean for students' career interest increases slightly from 14.21 (sd = 3.730) to 14.81 (sd = 3.745). However, the change is not significant ($p = 0.203 > 0.050$). The four items in the career interest constructs remain statistically insignificant as well. These unexpected results imply that there is little evidence to prove the correlation between the program and students' career choices in AI or computing science.

The unclear correlation can be partially attributed to the short timeframe. Students might not fully grasp the insights about AI or computing science, let alone consider these areas as future jobs. Even if they do, simply learning about a subject does not necessarily translate to job preference. More variables such as mentorship, duration of programming, and age should be considered. Additionally, a longitudinal study is needed to track the effects of similar learning programs on career interests over time.

6.3 Gender Disparity

Based on the between-subject effects no statistically significant gender difference is detected for self-efficacy ($p = 0.315 > 0.050$) or career interest ($p = 0.087 > 0.050$). Student outcomes in general are not differentiated by gender.

The undetected gender disparity in the sample can be attributed to the grade level of the sampled students. Students in Grades 3 to 5 might not have developed a gendered view of occupations, which research suggests may develop during Grades 6 to 8 [5]. Although it remains questionable whether gender will make a significant difference later in their lives, gender-inclusive programs are still crucial for underrepresented students at an early age before gendered views on career fields take root.

7 Limitations

Admittedly, the current study comes with shortcomings. The sample size is limited. A sample of 48 students with an unbalanced gender ratio may not accurately represent the demographics of Grades 3 to 5 students. A larger sample size with a more balanced gender ratio will provide a better representation of gender differences in the dependent variables.

The short time frame of the program also misrepresents students' actual learning outcomes. This project only took place over two class periods. However, elementary school students' career interests may take a much longer time to develop. Also, the scope of the knowledge in the curriculum is limited. More extensive work on the topic could be taught to elementary school students via long-term continuous programs and then evaluated.

Another limitation of this study is that the teacher is a strong female role model in computer science and AI. Thus, the survey may be vulnerable to observer bias. Students would speculate a "correct" response catering to the instructor's expectations. During the class periods, students might also tend to behave more extrovertedly than usual in front of their peers and instructors. These influences might cause students to report higher scores in their post-program surveys, resulting in overestimating the actual effects of the lesson.

Given the non-experimental nature of the study, there is no control groups for comparison. The lack of benchmarks may render questionable the actual effectiveness of the program. Nevertheless, whether to include a control group in this setting contains ethical concerns. Because the educational program pertains to students' prospects, it would be unfair to leave out a group of students for the sake of research observations.

8 Conclusions

The role of computing science and AI in society has increased dramatically, and educators need to prepare our young people for the digitalized future. A Fresh Squeeze on Data allows teachers to engage in interactive lessons with their students to discover concepts of bias in AI, machine learning, and other computing science fields. This study demonstrates how targeted interventions can affect students' self-efficacy and interest. However, it also demonstrates the need for longer term interventions and interventions conducted at earlier ages.

Future research can refine its methods in these four ways. First, longitudinal studies are needed to track the persistent effectiveness of the training program. Second, *A Fresh Squeeze on Data* can be integrated into other disciplines instead of being taught as a stand-alone subject. Thirdly, more demographic variables shall be considered, including race, socioeconomic class, school curricula, grade level, and ableness. Finally, similar studies should extend to areas where school affluence is differentiated.

Additional studies must also address questions such as the teachers' roles, parental support, and modeling [5]. Overall school support for the AI-related fields must be explored in greater precision. Nevertheless, by designing and implementing interventions such as *A Fresh Squeeze on Data*, teachers and schools may begin to close many of the learning gaps in AI, computer science, and data science that have persisted for far too long.

References

1. Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373. https://doi.org/10.1521/jscp.1986.4.3.359
2. Chai, C. S., Lin, P.-Y., Jong, M. S.-Y., Dai, Y., Chiu, T. K. F., & Qin, J. (2021). Perceptions of and Behavioral Intentions towards Learning Artificial Intelligence in Primary School Students. *Educational Technology & Society*, *24*(3), 89–101.
3. Condliffe, B., Quint, J., Visher, M. G., Bangser, M. R., Drohojowska, S., Saco, L., & Nelson, E. (2017). Project-based learning. A literature review. MDRC, 1-84.
4. Corbett, C., & Hill, C. (2015). *Solving the equation: The variables for women's success in engineering and computing*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036
5. Farland-Smith, D. (2009). Exploring middle school girls' science identities: Examining attitudes and perceptions of scientists when working "side-by-side" with scientists. *School Science and Mathematics*, *109*(7): 415–427. https://doi.org/10.1111/j.1949-8594.2009.tb17872.x
6. Fouad, N. A., & Santana, M. C. (2016). SCCT and underrepresented populations in STEM fields. *Journal of Career Assessment, 25*(1): 24–39. https://doi.org/10.1177/1069072716658324
7. Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1): 79–122. https://doi.org/10.1006/jvbe.1994.1027
8. Lent, R. W., & Brown, S. D. (1996). Social cognitive approach to career development: An overview. *Career Development Quarterly*, *44*(4), 310. https://doi-org.tc.idm.oclc.org/10.1002/j.2161-0045.1996.tb00448.x
9. Lin, P., Chai, C., Jong, M. S., Dai, Y., Guo, Y., Qin, J. (2021) Modeling the structural relationship among primary students' motivation to learn artificial intelligence, *Computers and Education: Artificial Intelligence*, 2, 1-7. https://doi.org/10.1016/j.caeai.2020.100006
10. ReadyAI (2021). *Data bias lesson plan*. Cloudera. https://freshsqueezekids.com/. August 18, 2022.
11. ReadyAI (2021). *A Fresh Squeeze on Data: Problem-Solving with Data.* ReadyAI LLC.