



Effects of Prior Academic Experience in Introductory Level Data Science Course

Noshaba Bhalli

nbhalli1@umbc.edu

University of Maryland Baltimore
County
MD, USA

Vandana Janeja *

vjaneja@umbc.edu

University of Maryland Baltimore
County
MD, USA

David Harding

harding@berkeley.edu

University of California, Berkeley
CA, USA

ABSTRACT

Data Science is an in-demand skill in the job market. To meet the demand, universities are offering data science courses or programs. These courses/programs not only give students data science skills, but also awareness of the field, increasing the likelihood of their opting for data science as a career. In this work we look at how the students' prior academic experience affects the outcome of an introductory data science class that requires no pre-requisites. We conducted the study in an undergraduate introductory data science class (IS 296) at the University of Maryland, Baltimore County (UMBC). The course was adapted from University of California Berkeley's Data 8 course. A pre and post survey was conducted to measure four factors in light of Social Cognitive Career Theory (SCCT): self-efficacy, identity, motivation and belonging uncertainty of students as a data scientist before and after taking the class. Our results show that although the course was designed requiring no pre-requisites, students with no prior programming experience and no statistics experience showed decrease in the four factors as compared to students with some prior experience.

ACM Reference Format:

Noshaba Bhalli, Vandana Janeja, and David Harding . 2024. Effects of Prior Academic Experience in Introductory Level Data Science Course . In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE 2024), March 20–23, 2024, Portland, OR, USA*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3626253.3635505>

1 INTRODUCTION

Increase in data generation, storage, and processing capabilities has increased the demand for data science related jobs over the past decade. US Bureau of Labor Statistics, predicted 36% growth in data science related jobs between 2021 and 2031 [1]. To meet this demand, universities have started offering introductory data science courses at the undergraduate level to expose students to the field. As these courses are new and in some cases adapted from other universities, there is a need to see the effectiveness of these courses across students from different backgrounds so that potential improvements can be made to create equal opportunities for all.

*Corresponding Author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0424-6/24/03.

<https://doi.org/10.1145/3626253.3635505>

In this paper we report the effectiveness of an introductory level data science course, IS 296 at the University of Maryland, Baltimore County (UMBC), adapted from UC Berkeley's Data 8 course requiring no pre-requisites. We consider prior academic experience in terms of: exposure to programming, exposure to statistics, and being a transfer student in the university. To measure the effectiveness of the offered course we look at **four factors** in light of social cognitive theory (SCCT): Self efficacy in data science, identity as a data scientist, belonging uncertainty in the data science setting, and motivation for data science (defined in section 2).

Motivation and Research Question The four factors of SCCT are indicative of individuals' beliefs about their ability to pursue and succeed in a career path. Looking at these factors can help us understand performance of students from different backgrounds and introducing interventions for the under performing groups in terms of the SCCT factors, such as making appropriate changes in curriculum or introducing a module for foundational elements at the beginning of the course. We investigate:

RQ: *"What are the effects of prior academic preparation on the four factors (self-efficacy, identity, motivation and belonging uncertainty) in an introductory level data science course with no pre-requisites?"*

2 METHODOLOGY

A pre-survey was given to IS 296 students at the start of semester and then the same survey was given at the end of the semester (post survey) to measure the change in self-efficacy, identity, motivation and belonging uncertainty. The survey anonymously recorded the demographics of students, to measure differences amongst students from different backgrounds.

Self-efficacy is individual's belief in their own ability to do something, measured by the following seven questions: i) I believe I will receive an excellent grade in this class. ii) I'm certain I can understand the most difficult material presented in the readings for this course. iii) I'm confident I can learn the basic concepts taught in this course. iv) I'm confident I can understand the most complex material presented by the instructor in this course. v) I'm confident I can do an excellent job on the assignments and tests in this course. vi) I'm certain I can master the skills being taught in this class. vii) I'm confident I can find the information I need to overcome programming challenges [2].

Data science identity describes how students think data science is related to who they are, measured by the following three questions: i) I see myself as a data science person. ii) I could imagine myself doing data science in the future. iii) In general, being good at data science is an important part of my self-image [3].

Belonging uncertainty measures the feeling of not belonging in

	Pre		Post	
	No	Yes	No	Yes
Total responses	17	14		
Prior programming experience	8	11	8	6
Prior statistics experience	8	11	6	8
Transfer student	7	10	5	9

Table 1: Distribution of students in the pre and post surveys

a setting, measured by following three questions: i) Sometimes I feel that I belong in data science, and sometimes I feel that I don't belong. ii) When something bad happens, I feel that maybe I don't belong in data science. iii) When something good happens, I feel that I really belong in data science [4]. The response option for all questions above was on scale of 1-7, 7 being strongly agree and 1 being strongly disagree. Belonging uncertainty questions were reverse scored to keep consistency in scoring of all four factors.

Students' Motivation describes students' interest to achieve a goal, measured by questions: For each of the following statements, please indicate how important it was in your decision to take a data science course: i) Data science meets a requirement for my major. ii) Data science skills are important to my major. iii) Data science skills will improve my chances of getting a good job after graduation. iv) Data science skills are important to my extra-curricular activities. v) Data science skills will help me make an impact and solve problems in society. vi) Data science is intellectually interesting to me. vii) Data literacy is important for everyone these days no matter what their career or major. The seven questions were on scale of 1-5, 1 being *Not at all important* and 5 being *Extremely important*.

3 RESULTS

We report the results of a pilot study done in Fall 2022 with 23 registered students (see Table 1 for survey responses by groups).

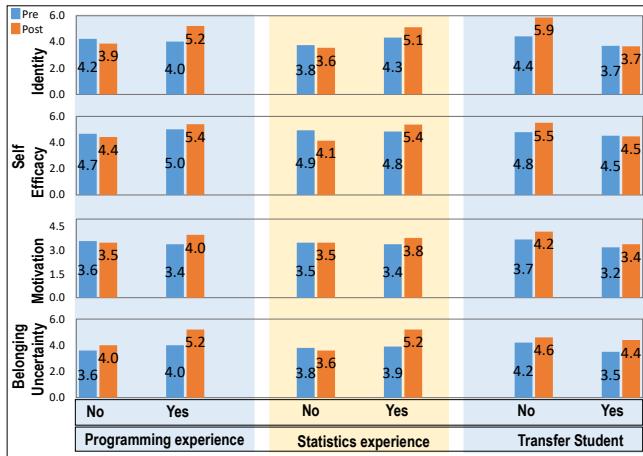


Figure 1: Pre and post survey comparison

We compared the pre and post survey responses to measure identity, self efficacy, motivation, and belonging uncertainty score with respect to prior experience in programming, prior experience in statistics, and transfer student status as shown in Figure 1.

Prior Programming Experience: Students with some prior programming experience showed improvement in identity, self efficacy, motivation, and belonging uncertainty (by 30%, 8%, 18%, and 30%

resp.). In contrast, students with no prior exposure to programming showed decreases in identity, self efficacy, and motivation (by -9%, -5%, and -3% resp.). Only belonging uncertainty improved in these students (by 11%). This decrease in majority of measures shows that programming experience is potentially necessary for the class.

Prior Experience in Statistics: Students with no exposure to statistics showed decreases in identity, self efficacy and belonging uncertainty (by -5%, -16%, and -5% resp.), the motivation measured the same. Where as students with prior experience in statistics showed improvement in all four measures. This makes prior knowledge in statistics potentially important for the students' interest. [5] reports similar findings based on a focus group study where IS 296 students without prior programming experience reported difficulty in programming assignments. All students in our survey with prior experience had it from taking course either in high school, community college or university and/or through self learning.

Transfer student: Previous work shows transfer students' performance decreasing in a new setting [6] as these students have limited/no prior experience in the new institution. We saw mixed trends for transfer vs non-transfer students. There is no negative change for the transfer students taking this class, however, all four factors are higher for non-transfer students than transfer students. Transfer students' identity and self-efficacy scored remained same, belonging uncertainty motivation showed positive improvement.

4 CONCLUSION AND FUTURE WORK

We outlined the effects of students' academic backgrounds on their motivation, identity, self efficacy, and belonging uncertainty in data science settings after taking an introductory data science class. Our results show that no prior experience in programming and/or statistics negatively impacts these four measures. Including introductory content at the beginning of the course, such as a week-long programming and statistics module, can potentially help students. Future work will look at data from additional semesters to develop a robust course plan supporting students from different backgrounds.

ACKNOWLEDGMENTS

This work is supported by NSF grant # 1915714

REFERENCES

- [1] U.S bureau of labor statistics (2022) url: <https://www.bls.gov/ooh/math/data-scientists.htm>, 2022. Accessed: 2023-09-06.
- [2] Paul R Pintrich et al. A manual for the use of the motivated strategies for learning questionnaire (MSLQ). 1991.
- [3] Gloriana Trujillo and Kimberly D Tanner. Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE—Life Sciences Education*, 13(1):6–15, 2014.
- [4] Gregory M Walton and Geoffrey L Cohen. A question of belonging: race, social fit, and achievement. *Journal of personality and social psychology*, 92(1):82, 2007.
- [5] Vandana Janeja, Maria Sanchez, Yi Xuan Khoo, Vacano Claudia, and Karen Chen. Adopting foundational data science curriculum with diverse institutional contexts. In *Proc. of the 55th ACM Technical Symposium on Computer Science Education*, 2024.
- [6] Philip E Carlan and Ferris R Byxbee. Community colleges under the microscope: An analysis of performance predictors for native and transfer students. *Community College Review*, 28(2):27–42, 2000.