EasyChair Preprint

№ 15867

How Students Learn Functions in an Integrated Introductory Data Science Module

Rabab Mohamed, Anupom Mondol, Dexiu Ma, Jie Chao, Kenia Wiedemann, Janet Bih Fofang, Barnas Monteith, Wanli Xing, Linlin Li, Yelee Jo, April Fleetwood, Lodi Lipien and Yuanlin Zhang

February 25, 2025

INDIVIDUAL PAPERS: How students learn functions in an integrated introductory data science module

Rabab Mohamed, Texas Tech University, rabab.mohamed@ttu.edu
Anupom Mondol, Texas Tech University, a.mondol@ttu.edu
Dexiu Ma, Texas Tech University, dema@ttu.edu
Jie Chao, Concord Consortium, jchao@concord.org
Kenia Wiedemann, Concord Consortium, kwiedemann@concord.org
Janet Bih Fofang, University of Maryland, bihjanetshufor@gmail.com
Barnas Monteith, Center for Science Engagement, barnas@engagescience.org
Wanli Xing, University of Florida, wanli.xing@coe.ufl.edu
Linlin Li, WestEd, lli@wested.org
Yelee Jo, WestEd, yjo@wested.org
April Fleetwood, Florida Virtual School, afleetwood@flvs.net
Lodi Lipien, Florida Virtual School, llipien@flvs.net
Yuanlin Zhang, Texas Tech University, y.zhang@ttu.edu

Session Description

According to NASEM (2018), data science has foundations in computing, mathematics, and statistics. However, at the K-12 level, these foundations are usually taught as standalone courses that are unconnected with each other. Students may struggle to see their connections. We proposed a framework unifying those foundations using mathematical logic. A core concept in mathematical logic is *function.* A general function has one or more possibly non-number inputs and an output. Data science motivates a comprehensive understanding of functions and provides extensive culturally relevant, real-world, and data-rich problems and applications for students to practice their understanding. It is interesting to know how well students understand functions. We developed a six-lesson online module with more than 100 in-lesson questions. Initial analysis of the students' answers to the questions shows that students can understand the basics of the general functions but have more difficulties in involved applications of functions.

Study students' learning of functions

Objective or purposes of the overall inquiry

Several data science curricula have been developed in recent years (DataScience4Everyone https://www.datascience4everyone.org/). Few studies have effectively unified core data science concepts. Our framework addresses this gap by exploring *functions* as a foundational concept in mathematics, statistics, and computer science, unified through mathematical logic. Functions at the secondary school level have garnered significant attention from researchers (e.g., Trujillo et al., 2023). However, this focus has been predominantly within the algebra context. There is a notable gap in research regarding students' understanding of functions in a broader context that includes computer science and statistics concepts. Our research investigates how well students understand *functions* through a data science module unifying its foundations in computing, math and statistics.

Perspective(s) or theoretical framework

Data science offers both practical motivations and real-world problem-solving opportunities, facilitating a comprehensive understanding of functions: from using functions to designing functions for *representing* information needed to solve given questions. Our framework for a unified data science course is rooted in Piaget's theoretical framework that positions *representations* in learning and sense-making to support the idea that active knowledge construction occurs through schemes structured for organizing and interpreting information (Piaget, 1972). This work explores approaches to use constructivist principles to connect students' experiences, learning interests, and understanding of real-life problems (e.g., Wilkerson & Polman, 2020; Bergin, 2016).

Methods of inquiry

Data Science Module and Data Collection. We developed a series of six 45-minute online lessons to cover various types of values (e.g, truth values -- true or false), functions on those values, general functions, and basic statistical concepts such as population and variables (through functions). Programming using Python and Jupyter Notebook is

seamlessly integrated into all the lessons. The following competencies on *functions* are fully covered: *identifying functions (iFun), reading functions (rFun)*, *writing functions (wFun)* and *designing functions (dFun)*. ***iFun*** is to identify functions (name, inputs and output). Given the *intended specification of a function* (i.e., a specifying a function using function name, inputs, output and how output is related to the inputs), **rFun** is to find the value of an expression using this and other (familiar) functions, and **wFun** is to write expression using this and other (familiar) functions to answer a question; **dFun** is to design a function for a concept, informal in daily life or in statistics, by writing its intended specification.

Participants and settings: 110 high school students are given a week to study these online lessons. Participants are selected on a volunteer basis, from online statistics and programming courses of a virtual school serving rural and racially diverse school districts.

Data Analysis. We have 140 in-lesson questions and collect all students' answers (about 12000). For each question, two researchers classified all the answers based on their similarities, and then discussed their classification and reached agreement. For each aspect of functions, we report the results on three types of typical questions: straightforward, in new context, and advanced. Questions are labeled Q1 to Q3 for each type. For iFun, students are tasked with identifying functions from arithmetic expressions using the mathematical operator addition (+) (Q1) and greater than (>) (Q2). For rFun, Q1 uses a function from the songs of a singer to their YouTube views while Q2 uses the concept *frequency* of statistics as a function. For wFun, Q1 uses the same function as Q1 of rFun, Q2 uses the *sort* function with input and output being vectors, and Q3 uses the *frequency* of statistics as a function. For dFun, Q2 uses the concept of *membership* in set theory as a function while Q3 uses the *(statistics) variable* as a function. For these questions, learners provided a total of 508 answers. We do not have all three types of questions for every aspect of the functions (see N/A in Table 1).

Results

The percentage of students giving fully correct answers for each question type of each aspect of functions is given in the table below.

	Identify functions (iFun)	Reading functions (rFun)	Writing functions (wFun)	Designing function (dFun)
Q1 (straightforward)	79%	88%	82%	N/A
Q2 (new context)	67%	61%	83%	20%
Q3 (advanced)	N/A	N/A	32%	11%

Table 1: Percentages of fully correct answers.

We next show some interesting information on students' understanding of functions. For Q1of iFun, all students are correct on the output for addition while 12% have problems with its inputs. For Q2 of iFun, 26% have difficulty on the output of the *greater than* function. For Q3 of wFun, 36% of students cannot use the function in a proper form, 21% know the form but missed one input. For Q2 of dFun, 68% is almost correct except for missing the input of a set or giving the wrong output description of the membership function. For Q3 of dFun, almost all students cannot use the proper form for designing a function.

Discussion and/or conclusions

Few existing works on data science (math or computer science) has covered general functions in a systematic way although they are a fundamental and unifying concept. Our data shows that the systematic treatment of general functions is within the reach of high school students (see Table 1). It is also clear that learning and practice are needed for students to master functions in more involved applications.

References

Bergin, D. A. (2016). Social influences on interest. Educational Psychologist, 51(1), 7-22.

data [n.d.]. DataScience4Everyone. https://www.datascience4everyone. org/.

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE—Life Sciences Education*, *18*(2), es2.

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664-672.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press.

Piaget, J. (1972). Development and learning. Reading in child behavior and development, 38-46.

Trujillo, M., Atarés, L., Canet, M. J., & Pérez-Pascual, M. A. (2023). Learning difficulties with the concept of function in maths: a literature review. *Education Sciences*, *13*(5), 495.

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. Journal of the Learning Sciences, 29(1), 1-10.