2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #40303

An Analysis of Data Analytics Curriculum Development through an NSF Research Experience for Teachers (RET) Program in Arkansas

Dr. Eric Specking, University of Arkansas

Dr. Eric A. Specking serves as the Assistant Dean for Enrollment Management and Retention for the College of Engineering at the University of Arkansas. Specking received a B.S. in Computer Engineering, a M.S. in Industrial Engineering, and a Ph.D. in Engineering from the University of Arkansas. His research interest includes decision quality, resilient design, set-based design, engineering and project management, and engineering education. During his time at the University of Arkansas, Eric has served as Principal Investigator, Co-Principal Investigator, or Senior Personnel on over 40 research projects totaling over $6.6 Million, which produced over 50 publications (journal articles, book chapters, conference proceedings, newsletters, and technical reports). He is an active member of the American Society for Engineering Education (ASEE) and International Council on Systems Engineering (INCOSE) where he has served in various leadership positions.

Dr. Shengfan Zhang, University of Arkansas

Shengfan Zhang is an Associate Professor in the Department of Industrial Engineering at the University of Arkansas. She received her Ph.D. and M.S. in Industrial Engineering from North Carolina State University. Zhang's current research focuses on developing methodologies and solution approaches in medical decision making, especially advancing predictive and prescriptive analytics for disease prevention and treatment. Zhang teaches courses on probability and statistics, predictive analytics, stochastic processes, quality engineering and management, simulation, etc.

An Analysis of Data Analytics Curriculum Development through an NSF Research Experience for Teachers (RET) Program in Arkansas

Abstract

The Arkansas Data Analytics Teacher Alliance (AR-DATA) Program, a Research Experience for Teachers (RET) Site, funded by the National Science Foundation (NSF) in its second year, has been promoting research-driven high school data analytics curriculum to reach underserved students in Arkansas. This work discusses the results from the first cohort of ten high school teachers in Arkansas who participated in a six-week summer program learning about data analytics, cutting-edge research in this field, and various engineering applications employing data analytics. They developed data analytics related modules for mathematics, computer science, and pre-engineering classes. In this paper, we first analyze the participating teachers' needs for module development and improvement, using information collected during the application process. We also summarize how data analytics related modules are incorporated in their current teaching materials. Through the analysis, we seek to explore how high school education in Arkansas is preparing students for next-generation workforce needs in analytics. In addition, we perform a descriptive statistics analysis of the learning modules created by the participating teachers through the AR-DATA program. We summarize the standards the teachers have used for their modules as well as the common ideas and topics of the learning modules. Through connecting the modules in different subject areas, we also analyze the possibilities of collaborative lesson plans that teachers in different fields can coordinate and teach together. Finally, we examine related topics in the post-secondary curriculum and propose how college professors and high school teachers can work together to strengthen education in data analytics to better prepare students for the workforce needs.

Introduction

Jobs with "data" in the title are increasing in popularity with industry shifting to data driven processes and decision-making enabled by new technology. Many universities across the United States are adding undergraduate and graduate degrees in data science or related fields to help fill these job demands, but kindergarten to twelfth grade education system is not keeping up [1]. Many schools focus on applications and resources, such as Microsoft Access and Excel. They tend to omit the ideas and theories in their lesson plans [2].

To help fix this problem, the University of Arkansas established in 2020 the state's first Research Experience for Teachers Site called the Arkansas Data Analytics Teacher Alliance (AR-DATA). AR-DATA aims to engage high school participants centered around data analytics through five components: 1) pre-program learning, 2) research activities with faculty and graduate students, 3) curriculum development, 4) curriculum implementation and testing, and 5) dissemination [3]. AR-DATA had its first cohort of 10 teachers in summer 2021. They attended a welcome week, which provided an overview of the program and an opportunity to learn about the data analytics research from all mentors. They then were paired with a mentor based upon their interest. They worked with a mentor and graduate students for 6 weeks to learn about data analytics and to create a module to implement in one of their classes. Throughout this time, each teacher met with curriculum experts to provide guidance and to ensure quality content aligned with standards. All participants presented the lesson plans at the end of the 6-week experience. They then piloted the lesson plans in their class throughout the year to edit and finalize based upon feedback from students to disseminate by posting on teachengineering.org and the AR-DATA website.

The rest of this paper is organized as follows. The next section provides demographic information and insights into the applicants and participants. In the third section, we use information from the RET application to provide insights into teachers' need for data analytics curriculum, expectations for our RET program, and current data analytics level of knowledge. The fourth section describes how participants created modules and provides a list with insights into their modules. We then connect the topics participants use in their curriculum to post-secondary industrial engineering course work. We end this paper with a conclusion and future work.

Applicants and Participants

This paper uses the applications and participants from the 2021 AR-DATA program to improve data analytics education in high school. The AR-DATA application contained 22 questions:

- Name;
- Ethnicity;
- Race;
- Gender;
- Veteran status;
- School and preferred email;
- Address;
- Phone number;
- School name;
- School district;
- Grade(s) taught;
- Subject(s) taught;
- Number of students taught;
- Need(s) for curricular improvements (needs);
- Expectations from the program (expectations);
- How will you disseminate what you've learned from this program;
- What do you know about data analytics; Do you teach it now (knowledge);
- Have you participated in a similar program;
- Agree to participate in all activities;
- Agree to participate in all program assessment surveys; and
- Where did you learn about this program.

In 2021, we received 20 applications and selected 10 participants. These applicants came from 13 school districts and 19 schools throughout Arkansas, seen in Figure 1. Figure 1 shows the locations of 2021 participants in red and the remaining applicants in blue.

Most applicants came from schools in Northwest or Central Arkansas. We gave priority to teachers from Northwest Arkansas since that was the grant's targeted geographic location. Table 1 summarizes the number of teachers from Northwest Arkansas. We extended teacher eligibility to outside of Northwest Arkansas due to COVID-19, which caused the program to shift to virtual for 2021. In 2021, 45% of applicants came from Northwest Arkansas, 60% of participants from Northwest Arkansas.

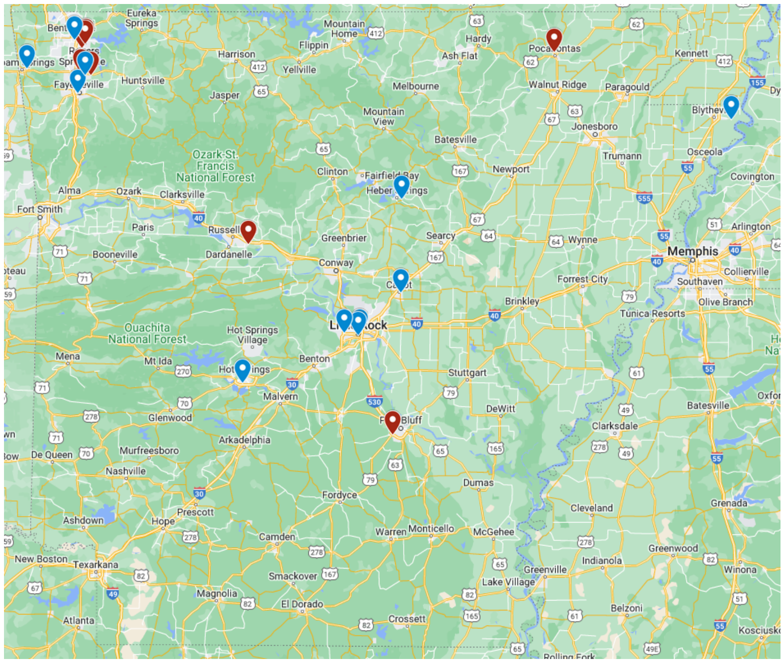

Figure 1. Location of Applicants and Participants

Table 1. Northwest Arkansas Applicant or Participants

NWA	Applicants	2021
No	11	4
Yes	9	6

Table 2 shows the number of applicants’ and participants’ schools that received funds for the rural and low-income status (RLIS) program. Many applicants came from schools not classified as rural or low-income, but 20% of AR-DATA participants came from schools who receive RLIS program funds.

Table 2. Rural Status of Applicants' or Participants' School

Rural	Application	2021
No	18	8
Yes	2	2

Next, we looked at the demographics of the applicants and participants. AR-DATA had an applicant pool that contained 45% of teachers self-identifying as female, while 60% of participants self-identified as female. Table 3 shows gender breakdown of applicants and participants.

Table 3. Gender Classification of Applicants and Participants

Gender	Application	2021
Female	9	6
Male	11	4

Seventy percent of applicants and participants self-identified as white. Table 4 shows the breakdown of applicants and participants by race.

Table 4. Race Classification of Applicants and Participants

Race	Application	2021
American Indian/Alaska Native	1	1
Asian, white	1	1
Black or African American	2	1
Prefer not to respond	2	0
White	14	7

In addition to gender and race, we looked at if applicants identified as Hispanic or previous serving in the military. Only one applicant self-identified as Hispanic and two identified as veterans. One of the veterans was selected as a participant.

After analyzing demographics, we looked at subjects taught by applicants and participants, summarized in Table 5. Thirty percent of applicants taught computers science, 40% taught math courses, and 30% identified as pre-engineering teachers. When looking further in the application prompts, we found that many of the pre-engineering teachers were not Project Lead the Way (PLTW) teachers but science or EAST teachers. We assume that this is because PLTW curriculum is standardized. Out of the 10 selected participants, 50% taught computer science courses, 30% taught math courses, and 20% taught pre-engineering courses.

Table 5. Academic Subjects Taught by Applicants and Participants

Subject	Application	2021
Computer Science	6	5
Math	8	3
Pre-engineering	6	2

Grade level taught was the final demographic information we analyzed. Forty-five percent of applicants claimed to have taught or currently taught all high school levels (e.g. 9^{th}, 10^{th}, 11^{th}, and 12^{th} grades). Table 6 summarizes grades taught by applicants and participants.

Table 6. Grades Taught by Applicants and Participants

Grades - All	Application	2021
10^{th}	1	0
10^{th}, 11^{th}, 12^{th}	3	1
11^{th}, 12^{th}	2	1

9th	5	4
9th, 10th, 11th, 12th	9	4

We did further analyses by looking at this same information broken by subject, gender, and race for all program years. It will be used for future work.

Applicant Needs, Expectations, and Current Knowledge

We asked three open response questions on the application to understand teachers' data analytics curriculum needs, expectations for the program, and current data analytics knowledge. This section analyzes those questions by using natural language processing (NLP) techniques.

We started by exporting the application data in a comma separated values (CSV) file. We then imported the data into an R data frame and cleaned up each data field. Next, we created sets of text files based upon participation year, rural status, and subject for needs, expectations, and knowledge. This created 27 text files and enabled us to look at needs, expectations, and knowledge from the lens of one classification. For example, we filtered the applicant data to include only 2021 participants and created three files (i.e. one for needs, one for expectations, and one for knowledge). After having the desired text files, we created a corpus by using the tm and SnowballC R libraries for text mining operations. This converted our files to lower case text, removed numbers, stop words, punctuation, and whitespace, and stemmed words to their base words (e.g. cleaning to clean). We used the corpuses to create term document matrices, which showed words and their frequency occurred. This enables us to easily create bar plots and word clouds to visualize the data. Additionally, we used the Syuzhet R library to perform sentiment analyses, which classifies text as positive or negative or as an emotion. We should note that not all words will have an emotion or positive/negative feeling connected to it. This enables us to get an understanding of the emotion written behind a response. This method is commonly used in product reviews. We created over 100-word clouds by using the term document matrices with the R library wordcloud.

We completed several comparisons based upon application versus participant, rural versus non-rural, and among the subjects teachers taught, but this section focuses on the 2021 applications. Focusing on applications gives us a larger sample to gain insights. We plan on publishing the additional comparisons in future work.

In general, all applicants and participants used more positive language compared to negative. This was consistent for need, expectations, and knowledge, seen in Table 7. We expected this since all applicants were applying for a professional development program.

Table 7. NCR Sentiment Analyses Results

	Needs		**Expectations**		**Knowledge**	
Scenario	**negative**	**positive**	**negative**	**positive**	**negative**	**positive**
All	17	126	6	124	15	64
2021	11	82	3	69	9	37

Many of the applicants used similar language when answering the three prompts (i.e. needs, expectations, and knowledge), such as "student”, “data”, “teach”, “program”, and “use”. These were some of the most frequently used terms when looking at a commonality cloud, seen in Figure 2, which shows words and their frequency that were included in all three files. We did not find any of the words surprising. Applicants used data 132 times, which was the most frequently used term. Figure 3 provides the frequency for the top 10 words used by applicants in all three questions. For example, one teacher stated, “I am hoping this experience will help me show my students, how mathematics relates to the real world…My expectation is to help my students learn about the field of data analytics and the careers that go with it.”

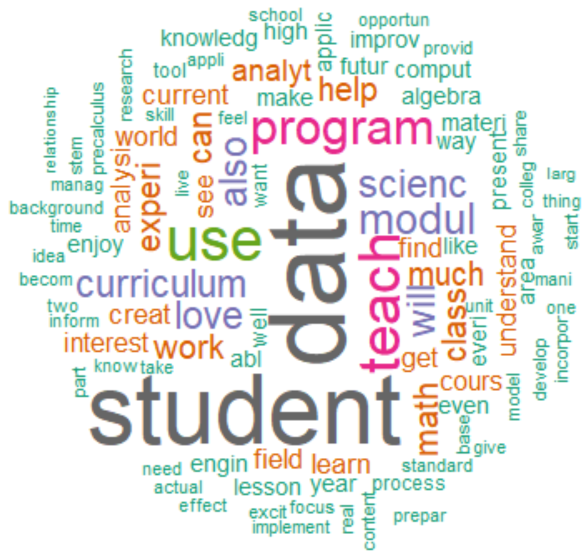

Figure 2. Commonality Cloud for Needs, Expectations, and Knowledge of Applications

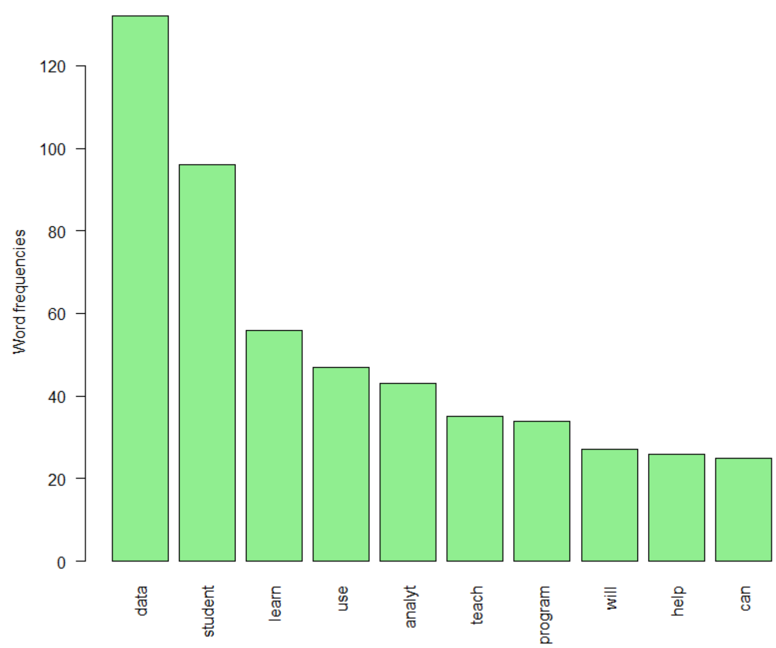

Figure 3. Applicant Needs, Expectations, and Knowledge Commonality Cloud Term Frequency

We gain additional insights when looking at a comparison cloud, which compares frequency of terms for multiple documents and plots the word where it occurs the most. Figure 4 show the comparison cloud for applicants' needs (lower right and in purple), expectations (upper right and in green), and knowledge (left and in orange). Darker and larger words indicate occurring more frequently, while smaller and lighter colors indicate occurring less frequently. For example, "data" is the largest word, in the center, and colored orange. This means that "data" occurred the most and in all three of the responses. The comparison cloud code classified it as Knowledge (orange) since it occurred the most in the knowledge text responses.

Figure 4. Comparison Cloud for Needs, Expectations, and Knowledge of Applications

Figure 4 shows that 2021 applicants used words, such as "need" and "comput" when describing their needs. Teachers often referenced computer science or computer, which caused "comput" to have a high frequency. One teacher stated, "I have enjoyed teaching Computer Science but I know with the fast pace and changing times that everyone can use more knowledge, I feel that I need help creating lessons with Data that is used in the world today." Most teachers wanted their students to have better experiences collecting and using data to solve real world problems through experiential learning. One participant stated when answering the need question, "My current needs include implementing research and connecting activities to real life experiences…students get to 9th grade and still struggle to use and read graphs…love to find new and exciting ways to collect, analyze, and interpret data."

Applicants used words, such as “learn”, “expect”, “profession”, “understand, and “hope” when describing their expectations. Many applicants discussed creating relationships with other teachers and developing new skills. One applicant said, “Not only would this give me a chance to continue my learning about the field that I'm passionate about, but it allows me to learn about the many careers paths that require a science/math/computer science background.” Another stated, I hope to gain a deeper understanding as to how to bring critical thinking skills to students that will work with data day in and day out……gain a deeper understanding of data that we encounter every day, and how we can present that information to the students.”

The most used word, “data”, appeared the most in the knowledge question. Many teachers, when talking about their knowledge used phrases, such as “data analytics”, “analyzing data”, and “gathering data”. It should be noted that teachers used it to state what they did and did not know. One teacher stated, “I understand that data analytics is the process of analyzing raw data to find trends and answer questions….[we] do not have well-defined data analytics modules.” In general, applicants stated that they knew statistics and had previous experience with data analytics, but it was evident that there is a need for data related teaching materials. For example, an applicant stated at the end of the knowledge response, “I am excited to learn more and integrate more data related modules into my teaching materials.” One teacher did state, “I have zero knowledge of what this [data analytics] is. It looks interesting, so why not?”

Participant Modules

An overview of the modules developed by the 2021 RET participating teachers is presented in Table 8 with information on module name, targeted class (grade, subject, and duration of modules), data analytics related methods and tools covered in the lesson plans, engineering and computer science application areas, and targeted curriculum standards. In the 2021-2022 program, there are a total of nine lesson plans developed, with two teachers combining their plans into a full-day (or over a week) lessons and activities. All lesson plans have a duration of 3-5 class periods, with 45-50 minute per period. Some also require activities outside class period for data collection. Among the nine lesson plans, three targeted mathematics classes, three for computer science classes, two for science/pre-engineering classes (chemistry, biology), and one can be used in both mathematics and computer science classes.

All lesson plans have a data component, with students either collecting data on their own, or teachers providing data for them to observe. All lesson plans focus on how to analyze the data using data visualization (e.g., creating scatterplot) and simple statistics or mathematics calculation (e.g., correlation, matrix operation). Most lesson plans allow students to explore descriptive and predictive statistics methods, such as fitting lines, and some predictive or classification models. Students also learn how to implement their data cleaning, visualization, or simple predictive analytics using spreadsheet modeling and some programming (mostly for computer science classes). All lesson plans are developed according to curriculum standards, e.g., Next Generation Science Standards (NGSS), Arkansas Mathematics and Computer Science Standards. Three teachers did not specify detailed standards targeted for their lesson plans. They will have to specify them in their final product. There are various application areas associated with the lesson plans. For example, two lesson plans focus on transportation, one is using videos online to observe data related to traffic (e.g., number of cars, cars turning), and the other collects their own traffic data near their school and use simulation to compare and present engineering solutions (i.e., adding traffic light, roundabout, or keep four-way stop signs). Other application areas include energy (electrical engineering), DNA tiles (computer science), thermodynamics (mechanical engineering), environment or civil engineering (two plans), fracture mechanics (mechanical and civil engineering), image recognition and classification (industry engineering and computer science). The lesson plans are also mostly aligned with their faculty mentor expertise areas.

Table 8 Summary of Developed Lesson Plans

Module Name	Targeted Class	Analytics Methods/Tools	Application Area	Standards
Data Analytics in the Energy Sector	8th Grade Math (Algebra III) 4 Class Periods	Data collection, Data visualization, Spreadsheet modeling, Statistics (e.g., correlation)	Energy, Battery	Arkansas Mathematics Standards (8.SP.A.1, 8.SP.A.2)
An Introduction to Data Analytics Using Abstract Tile Assembly Models (aTAM)	9th-12th Computer Science 3 Class Periods	Binary and number systems, Data representation and visualization, Data analytics	DNA Tiles (biology, nanotechnology)	Arkansas State Standards (Not Specified)
Modeling Thermodynamics with Machine Learning	10th-12th Grade Chemistry 3-5 Class Periods	Predictive modeling (machine learning), Spreadsheet modeling	Physics, Thermodynamics, Energy	Next Generation Science Standards (NGSS HS-PS3-1, HS-PS3-4, HS-ETS1.1—1.4)
Analyzing Traffic in an Algebra Class	9th – 12th Grade Math 3 Class Periods	Data types, Data cleaning, analysis, and visualization, Data fitting (line fitting)	Transportation	Arkansas Mathematics Standards (Not Specified)
Water pH and Quality Survey	9th Grade Biology, Chemistry 3-4 Class Periods	Data collection, Data visualization, Data fitting, Spreadsheet modeling	Environmental Engineering	Arkansas NGSS (PSI6-ETS1-2, Common Core RST.11-12.9)
Stretch It Out	11th-12th Precalculus, Trigonometry 5 Class Periods	Data collection, Mechanistic modeling	Fracture Mechanics	NGSS (HS-ETS1-2), CCSS (Math.Content.HSF. TFA.3, TF.C.9), Arkansas Precalculus (Grade 9-12: T.3.PC3, PC5)
Lessons Based on a Bystander Anonymization Article	9th-12th Computer Science, Algebra II, III, Precalculus, 4-5 Class Periods	Qualitative analysis, Data visualization, Matrices/Vectors, Image Processing, Deep Learning	Image Recognition	Computer and Mathematics State Standards (Not Specified)
Traffic Stop Data Analysis	9th EAST Program, Computer Science 3 Class Periods	Data collection, Simulation, Spreadsheet modeling, Data Visualization	Transportation	NGSS (HS-ETS1-2—1-4)
Data Day – A Daylong Introduction to Data and Data	9th Computer Science	Machine learning, Data collection, Data visualization,	Image Classification, Environmental Engineering	NGSS (HS-ETS1-2), Arkansas Computer Science Standards (CSPG.Y.1.3.1-5,

Analytics (Two Teachers)	A full day or over a week “data bootcamp”	Classification, Engineering design		1.7.2), Arkansas Algebra 1 Standards (HSS.ID.B.6 22, A.1-A.3 22, B.5 22)

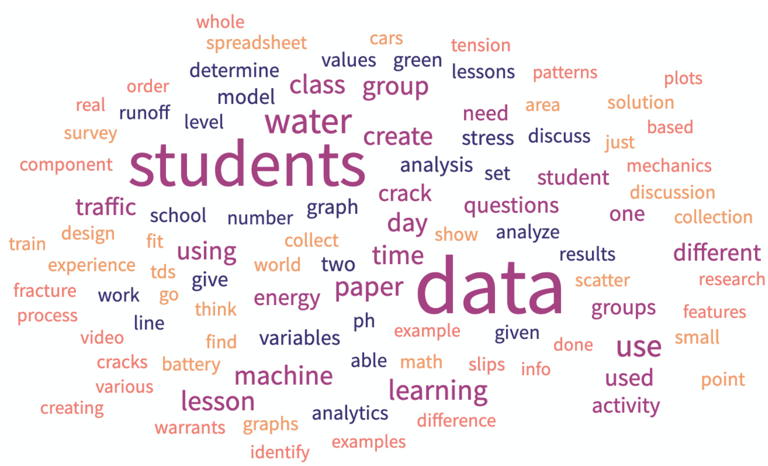

Figure 5. Frequency Cloud for Lesson Plans Developed by 2021 RET Participants

We combined all lesson plans and built a word cloud to show commonality among developed plans. Not surprisingly, “data” and “students” are the two most frequently appearing words in the lesson plans. Other words shown also demonstrate different focus areas of the lesson plans, e.g., water (environmental engineering), traffic (transportation), crack (fracture mechanics). The word cloud demonstrates the methods to cover, e.g, machine learning, scatter plots. The delivery of contents mostly includes working on paper, in group activities or discussions, spreadsheet modeling, etc.

There is some limitation to this approach. Participants did not develop the lesson plans in the same format. While we encourage teachers to develop plans using the TeachEngineering format, they can also use other lesson plan template to create their modules. Some teachers have created very detailed plans for their lessons, such as list of questions and list of steps for activities, while other teachers have an overview and day-by-day summary for their lessons.

Connecting K12 to Industrial Engineering

We compared the analytics methods and tools presented in the third column of Table 8 with Industrial Engineering curriculum at the University of Arkansas. Table 9 summarizes the key topics shown in both the AR-DATA lesson plans and the Industrial Engineering catalog course description.

Table 9 Comparing K-12 Data Analytics Topics with Industrial Engineering Curriculum

Data Analytics K-12 Topic	**Industrial Engineering Courses Covering the Topic**
Data collection (5) and data visualization (5)	INEG 2103 Introduction to Industrial Engineering
Spreadsheet modeling (4)	INEG 2103 Introduction to Industrial Engineering INEG 4683 Decision Support in Industrial Engineering

Data summary and representation (1)	INEG 2223 Computing Methods for Industrial Engineers II INEG 2314 Statistics for Industrial Engineers
Statistics (e.g., correlation) (1), regression (Data fitting) (2)	INEG 2314 Statistics for Industrial Engineers I INEG 4143 Data Mining
Machine learning (2), deep learning (1)	INEG 3333 Statistics for Industrial Engineering II INEG 4143 Data Mining INEG 4163 Introduction to Modern Statistical Techniques for Industrial Applications

Each number in the parentheses in the first column of Table 9 corresponds to the number of lesson plans (out of nine) that reflected the specific data analytics topic. It can be seen that data collection, data visualization, and spreadsheet modeling are the common topics chosen, and they are well-connected to industrial engineering curriculum. It is not a surprise that the related college courses are mainly in statistics, basic modeling and computing classes, and advanced classes related to data analytics.

Conclusion and Future Work

In this paper, we presented a descriptive statistics analysis of the learning modules created by the participating teachers through the AR-DATA program. We summarized the standards the teachers have used for their lesson plans as well as the common ideas and topics of the learning modules. It can be seen that modules that were developed for different subject areas can share the same methods or tools in data analytics. It is possible that teachers in different subject areas can collaborate and develop connecting modules to further engage students while leveraging resources and learning opportunities. We also examined related topics in the college curriculum and found a strong connection between the developed lesson plans and industrial engineering courses. As a future research direction, when we have multiple years of data from the program, we can draw a better inference on connections among different subject areas and classes, and propose more detailed ideas on curriculum collaboration in data analytics for the K-12 environment.

Acknowledgement

This work is supported by the AR-DATA RET Program funded by the National Science Foundation (NSF: #1953733).

References

1. PwC and Business-Higher Education Forum. “Investing in America’s data science and analytics talent,” PwC and Business-Higher Education Forum, 2017
2. Aqlan, Faisal, Joshua C. Nwokeji, and Abdulrahman Shamsan. "Teaching an Introductory Data Analytics Course Using Microsoft Access® and Excel®." In *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1-10. IEEE, 2020.
3. Zhang, S., Specking, E., Alimohammadi, M., Boykin, A., Bell, S., Schubert, K., Davis, S. “Establishing a Research Experience for Teachers Site to Enhance Data Analytics Curriculum in Secondary STEM Education.” In *Proceedings of the 2021 ASEE Midwest Conference*, 2021