What Is Interesting and Relevant About Cybersecurity? - NLP Analysis of a Survey of CS Students

Cheryl Resch
Engineering Education Department
University of Florida
Gainesville, FL, USA
cheryl.resch@ufl.edu
0000-0003-1668-8850

Jinnie Shin
School of Educational Development and Organizational Studies
University of Florida
Gainesville, FL, USA
jinnie.shin@coe.ufl.edu
0000-0002-1012-0220

Christina Gardner-McCune
Computer and Information Science and Engineering
University of Florida
Gainesville, FL, USA
gmccune@ufl.edu
0000-0002-4397-9162

***Abstract*—Cyber attacks are a common feature of current news and many of them are the result of easy to avoid vulnerabilities in software. It is imperative that students graduating from an undergraduate Computer Science (CS) curriculum understand the consequences of vulnerable code. When developing lessons and assignments, it would be useful to have a sense of students' attitude toward cybersecurity and appreciation of the need to write secure code. This paper describes an analysis of the results of a survey of students in core CS courses at our large public university, in which students answer free response questions about what they find interesting and relevant about cybersecurity. The survey was conducted in Fall 2022 and repeated in Spring 2023 after cybersecurity interventions were introduced into several core CS courses. We performed a Natural Language Processing (NLP) analysis of the free response answers to determine the overarching themes in the responses. We found that the most prevalent topics students are interested in are cryptography and penetration testing, and did not change over the two semesters. In answer to the question about the relevance of studying cybersecurity, we found that as students progress through the curriculum, what students find relevant moves from protecting their personal data to its importance in job duties and writing secure programs. When developing lessons and assignments, it may be helpful to introduce cryptography or penetration testing to engage students. Also, students should be taught early and often about the relevance of cybersecurity in their future job duties.**

Keywords—cybersecurity education, natural language processing

I. INTRODUCTION

Data breaches caused by software vulnerabilities in commercial products can lead to interruptions in public services, monetary loss, and loss of privacy. The 2022 Verizon Data Breach Investigation Report [1] indicates that there were 5,212 data breaches in 2022 in the United States and the number of breaches increases every year. Despite increased tracking and abatement of software vulnerabilities, Gueye and Mell [2] report that the most prevalent software errors have not changed much since vulnerabilities were first cataloged. Indeed, MITRE [3] lists the top four most dangerous software vulnerabilities of 2022 as:

1) Out-of-bounds write (buffer overflow)
2) Improper Neutralization of Input on Webpage Generation (cross-site scripting)
3) Improper Neutralization of Special Elements used in a SQL Command ('SQL Injection')
4) Improper input validation

Software vulnerabilities can be reduced or eliminated when developers use principles of secure programming. The Association for Computing Machinery (ACM) included the Information Assurance and Security knowledge area in its Computer Science curriculum starting in 2008 [4]. The 2008 Computer Science curriculum included Secure Programming, and the 2013 Computer Science curriculum updated the knowledge area to also include Defensive Programming and Secure Software Engineering [4] [5].

When determining how to distribute secure programming topics in the Computer Science curriculum, it would be useful to have a sense of students' knowledge of and attitudes toward cybersecurity when they enter the curriculum, and how it changes as they progress through the curriculum. This work analyzes the responses to a cybersecurity survey taken by CS students in different core courses of a CS curriculum. Students were asked what they find interesting and relevant about cybersecurity. NLP analysis with unsupervised machine learning approaches were used to effectively and efficiently analyze answers to free response questions.

In this paper we explore the following research questions:

- RQ1 – What do CS students find interesting and relevant about studying cybersecurity?
- RQ2 – How do CS students' attitudes differ according to where they are in the curriculum?
- RQ3 – Do CS students' attitudes differ if they have been exposed to cybersecurity learning interventions?

II. PRIOR WORKS

Previous literature has surveyed college students on their knowledge of and attitudes towards cybersecurity [6] [7] [8] [9] [10] [11]. This previous work has focused on cybersecurity awareness of college students in general. The work described in this paper can be distinguished from this previous work because our work surveys Computer Science students specifically, and includes the ideas of writing secure software rather than only cybersecurity awareness. This work also correlates the knowledge and attitudes toward cybersecurity with where students are in the cybersecurity curriculum.

More recently, English and Maguire [12] surveyed 4th year CS students taking a computer security course and master's level students taking a business-focused cybersecurity course. They asked similar questions to those of this work, including "What do you want to learn about cybersecurity", and "How do you perceive knowledge of cybersecurity contributing to your future professional role?" They received 126 responses. 73% responded that knowledge of cybersecurity was very important to their future professional role. Topics students wanted to study included cryptography, ethical hacking, writing secure code, and web security. Our work extends this to survey students throughout the CS curriculum, and conducts the survey more than once to explore changes in attitudes over multiple semesters.

NLP has been used to gather insight on large corpora relating to cybersecurity, and student responses to reflection prompts. In the cybersecurity domain, NLP has been used to mine Twitter and vulnerability reports in order to quickly gain insight on important cybersecurity related events. Georgescu [13] describes using NLP for the development of a cybersecurity vocabulary. The author developed categories such as attacker, vulnerability, and software, and classified words in cybersecurity texts according to their category. Roy et al. [14] use Word2Vec with continuous bag of words (CBOW) and skip-grams to develop document embeddings for descriptions of malware. They developed categories such as vendor, operating system, and vulnerabilities. Gasmi et al. [15] use a Long Short-Term Memory (LSTM) neural network to perform named entity recognition (NER) for documents from the National Vulnerability Database (NVD). These works provide insight into categorizing cybersecurity words. The words used by students will not be at the level of detail described in [13], [14] and [15]; students will use general words to describe cybersecurity. But the categories developed provide insight as we study cybersecurity topics in student responses.

Now we review work done in analyzing short student reflections using NLP. DeLin et al. [16] compare the use of Latent Dirichlet Allocation (LDA), Gibbs Sampling Dirichlet Mixture Model (GSDMM) and Word2Vec clustering for finding topics in responses to reflection prompts asking students to identify the most challenging topics in a course. They found that Word2Vec embeddings and K-means clustering performed the best at matching the ground truth human labeling of reflections with topics. Dorodchi et al. [17] explore the use of clustering and sorting of student reflections in a Software Engineering class. Similar to [16], the reflections are intended to provide insight to instructors on what students are finding challenging in the class. They used RoBERTa to generate sentence embeddings and K-means clustering to find clusters. They found that sorting the reflections by clusters helped instructors more quickly gain insight into what students needed help with.

This work expands this previous work to explore short student reflections on cybersecurity to extract topics. Specifically, we adopted state-of-the-art sentence embedding techniques with a strategy similar to that found in [16] and [17] to find the prevalence of cybersecurity topics in the student responses.

III. STUDY CONTEXT

In this paper we analyze the responses of undergraduate students in five core Computer Science (CS) courses at a large R1 university to a survey of cybersecurity knowledge and attitudes. The five courses are: Programming Fundamentals 1 (CS1), Programming Fundamentals 2 (CS2), Advanced Programming Fundamentals (CSA), Computer Organization (CompOrg), and Operating Systems (OS). CS1 and CS2 are the first two courses of the CS curriculum. CS1 is a prerequisite of CS2. CSA is a one semester course that covers material in CS1 and CS2. Students place into this course if they received a 4 or 5 on the Advanced Placement (AP) Computer Science A Exam [18]. CSA is only offered in fall semesters at our university. CS2 or CSA is a prerequisite of CompOrg, and CompOrg is a prerequisite of OS. Thus, a student can be in only one of the surveyed courses. The survey was optional and students received extra credit for participating. The study has been approved by the Institutional Review Board (IRB) at our university. The survey was offered in the first month of the Fall 2022 semester and the first month of the Spring 2023 semester.

Table I provides the enrollment in the courses from which participants were recruited and the number of participants from each course in each semester.

TABLE I. SURVEY PARTICIPATION

Course	F22 Enrollment	F22 Survey Participation	S23 Enrollment	23 Survey Participation
CS1	771	440	550	494
CS2	403	379	627	326
CSA	132	118		
CompOrg	678	640	455	388
OS	350	196	389	254

In addition to the survey, assignments focused on input validation, buffer overflow, and cross-site scripting were

introduced into the CS1, CompOrg, and Software Engineering courses at our university in Fall 2022, after students completed the survey. Software Engineering is a course that may be taken at the same time students are taking CompOrg or Operating Systems. Thus, in Spring 2023, students in CS2 are likely to have had an assignment on input validation in the previous semester, and students in Operating Systems may have had an assignment on cross site scripting or buffer overflow in the previous semester.

IV. METHODS

Responses are collected from the survey, and the course the participant is taking is saved with the response. Table II shows sample responses and courses. Embeddings were calculated for each response using the pre-trained Masked and Permuted Pre-training for Language Understanding (MPNet) [19] sentence embedding model. The model uses Masked Language Modeling (MLM) and Permuted Language Modeling (PLM). MLM learns the context of words in sentences by randomly masking words and training the model to predict the missing words. PLM learns the significance of the order of words. The MPNet transformer takes a sentence as input, and produces a vector of numbers that describe the sentence. The Hugging Face [20] pre-trained model was used to create sentence embedding vectors. This model has been trained on over one billion English language documents.

After the sentence embedding vectors are created, the DBScan algorithm [21] [22] is used to create clusters. DBScan is a widely used density-based clustering algorithm, and appropriate for this application because the number of clusters does not have to be specified before running the algorithm, and the algorithm can find arbitrarily shaped clusters.

After the DBScan algorithm is run to find clusters, the clusters are analyzed to discover overarching themes in the data. The first author analyzed the clusters and postulated overarching themes. The first author and three additional coders matched clusters to overarching themes and identified miscategorized responses in the clusters. Table II shows part of a sample cluster. The theme of this cluster is "white hat hacking". A coder reads the statements and identifies an overarching theme and identifies statements that don't match the theme of the cluster. The italics statement is an example from this cluster that does not match the theme.

TABLE II. SAMPLE "WHAT DO YOU FIND INTERESTING ABOUT CYBERSECURITY" CLUSTER

Response	Course
I always thought that "white hat" hacking was cool as it was basically getting paid by a company to hack into their database	CS1
I am interested to learn how white-hat hackers use their skills to help companies improve their cybersecurity. I am also interested in how hacking and cybersecurity applies to forensic evaluations	CS2
White hat hacking because it allows you to think from the perspective of the hacker but you are committing similar actions in an ethical manner	CSA
I find hacking and encrypting messages interesting	CompOrg
I find hacking and whitehat hacking interesting	CompOrg
I find white-hat hacking to be a very interesting field and a fascinating instance as to how hackers actually exist and are employed	CompOrg

Inter-rater reliability between the first author and three other coders was calculated using Cohen's kappa [23]. For the inter-rater reliability calculation, three clusters were given to each coder (a total of nine clusters). The first author and the coder each identified the overarching theme and the number of statements that do not match the theme. For all nine clusters, the first author and the coder agreed on the overarching theme of the cluster. For agreement on miscategorized responses, Cohen's kappa [23] values for the coders are 0.645, 0.706, and 0.735, which is interpreted as significant agreement. Next, for each course, we calculated the proportion of students with a response in each overarching theme. We then compared proportions for the different courses or semesters. For example, we compared the proportion of CS1 students in Fall 2022 who indicated they found cryptography interesting to the proportion of CS2 students in Fall 2022 who indicated that they found cryptography interesting. To test whether two proportions are significantly different, we calculate the non-parametric parameter Z [24]

$$Z = \frac{(\rho_1 - \rho_2)}{\sqrt{(1/\eta_1 - 1/\eta_2) \times \rho_o \times (1 - \rho_o)}} \quad (1)$$

where η_1 is total the number of responses in one group (e.g. CS1 students in Fall 2022), η_2 is the total number of responses in the second group (e.g CS2 students in Fall 2022), ρ_1 is the proportion of responses in the first group in a particular category (e.g. proportion of CS1 students in Fall 2022 who responded "cryptography"), ρ_2 is the proportion of responses the second group in the particular category, and ρ_o is the overall proportion of in the particular category. A Z-value with an absolute value above 1.94 indicates that the two tailed probability that these proportion values are from the same population is less than 0.05.

V. RESULTS

A. What do you Find Interesting

There were a total of 1560 responses in Fall 2022 to the question "What do you find interesting about cybersecurity". Fig. 1 shows the most prevalent themes and the fraction of students in each course with a response in that theme. For example, there were 440 responses from students in CS1, and

29% of them were in the theme "nothing/I do not find it interesting". The most prevalent themes (besides no answer, nothing, or vague answers) were cryptography (9.7%), white hat hacking (8.3%), keeping data safe (7.8%), and how hacking works (4.7%). White hat hacking, penetration testing, and finding vulnerabilities are all terms that describe the same thing: finding vulnerabilities in an enterprise for the purpose of shoring up defenses and protecting it, so clusters with these terms were combined into one theme.

Smaller clusters had the following themes:

- Adapt/evolve/cat and mouse/arms race
- Firewalls / network security
- Writing secure software / securing apps
- Phishing
- Cross-site scripting/sql injection
- Exploiting vulnerabilities
- Cloud security
- Block chain
- Reverse engineering
- Buffer overflow
- Distributed denial of service
- Hardware security
- Hacking subculture
- Social engineering

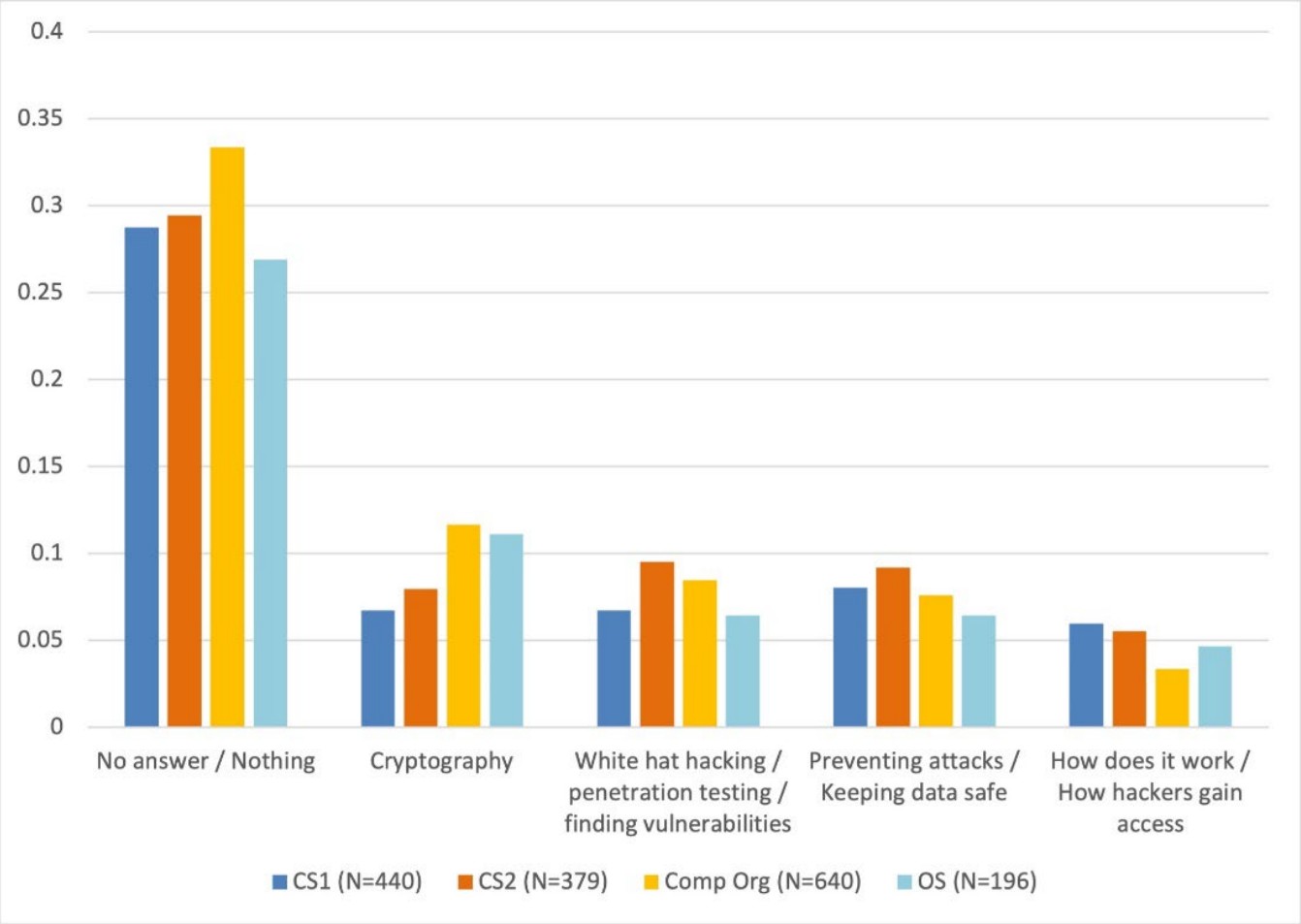

Fig. 1. Fall 2022 "What Do You Find Interesting" Results

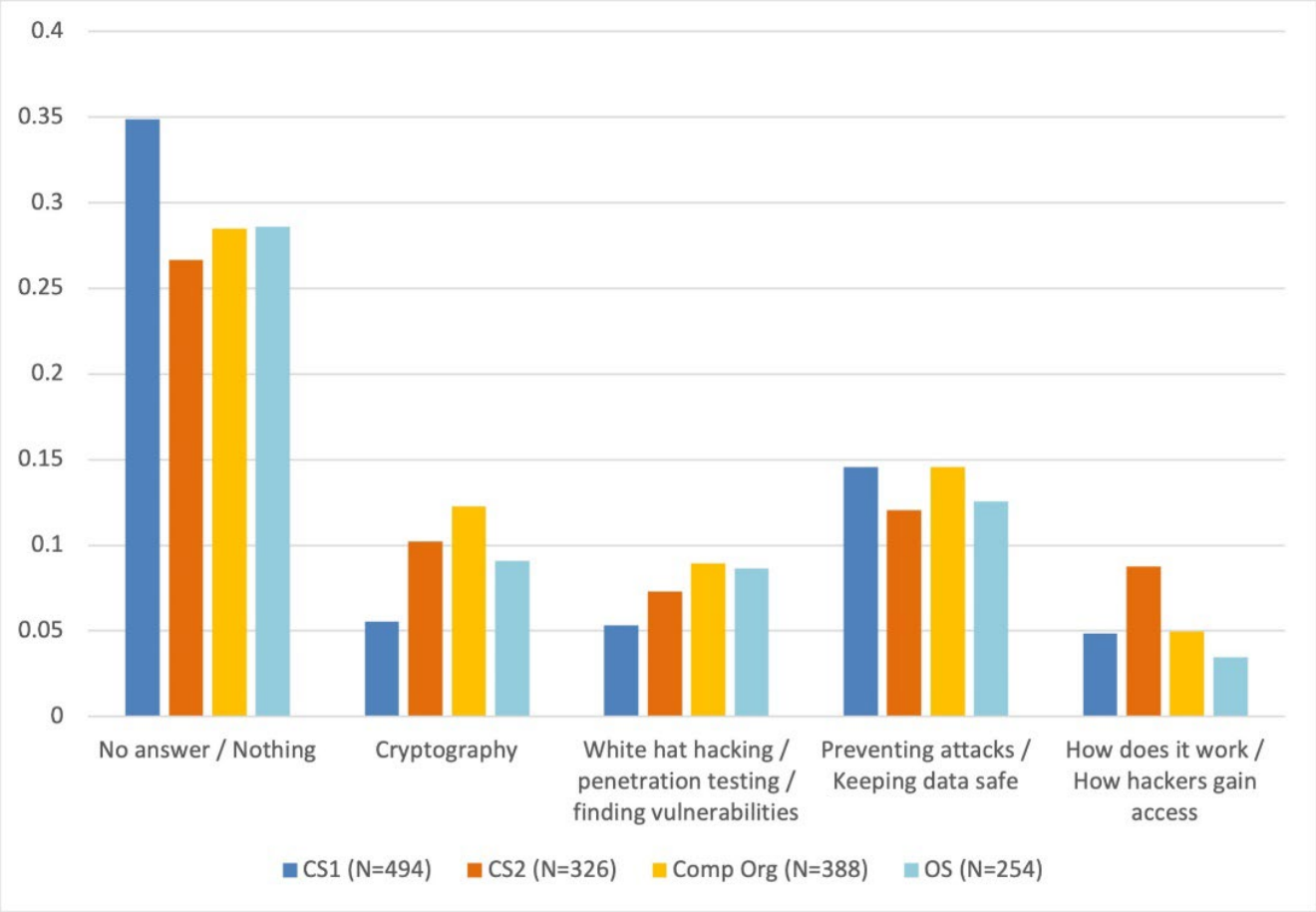

Fig. 2. Spring 2023 "What Do You Find Interesting" Results

We found that 7.9% of the responses had a theme that did not match the theme of the rest of the responses in the cluster. Also there were several clusters, representing 130 responses, for which coders could not find a common theme in the responses that were placed in the cluster. This data was excluded from the analysis.

Fig. 2 shows the fraction of students in each course with a response in the most prevalent themes for the Spring 2023 survey. There were a total of 1225 responses. We found that 9.6% of the responses had a theme that did not match that of the other responses in the cluster. Similar to the 130 responses in Fall 2022, in Spring 2023 there were several clusters, representing 64 responses, for which coders could not find a common theme in the responses that were placed in the cluster. This data was excluded from the analysis.

The most prevalent themes (besides no answer, nothing, or vague answers) were protecting data (13.8%), cryptography (9.0%), white hat hacking (7.3%), how hacking works (5.6%), and personal privacy (4.0%). The proportion of answers in "keeping data safe" was higher in Spring 2023 compared to Fall 2022, and "privacy / personal security" was a new prevalent theme in Spring 2023.

We found two significant differences across courses in the same semester. The first was in Fall 2022, where we identified a significantly higher proportion of CSA students who responded that they were interested in cryptography and white hat hacking compared to students in CS1 (Z=1.94). The second significant difference across courses was identified in Spring 23, where a significantly higher proportion of CompOrg students responded that they were interested in cryptography (Z=3.24) compared to students in CS1. When comparing categories across semesters, Fall 2022 to Spring 2023, the only significant change identified was for the "keeping data safe" category, where the proportion of students in all courses was significantly higher in Spring 23 (13.8%) compared to Fall 22 (7.8%), with a Z parameter value of 4.87.

B. How is Studying Cybersecurity Relevant to Your Future?

For the responses to the question "How is Studying Cyber-security Relevant to Your Future," clusters were grouped into the overall themes described in Table III. 16.1% of the Fall 2022 responses and 13.6% of Spring 2023 responses did not match the overall theme of the cluster.

TABLE III. THEMES OF "HOW IS LEARNING ABOUT CYBERSECURITY RELEVANT TO YOUR FUTURE" RESPONSES

Theme	Description	Sample Response
Write secure software	Mentions theme of writing software or code securely or that what they produce is safe	I think it would be I plan to become a software engineer so when building anything of the sort I would always have to make sure the user cannot abuse and exploit my code to obtain sensitive information
Use in my job or major	Mentions that it is important in job duties or chosen field but no mention of safe code or products	Cybersecurity will be important for whatever field I enter especially if I enter a field such as data science or software development. As the world relies more and more on technology cybersecurity becomes more of a necessary area to learn
Job opportunities or possible career path	Knowing cybersecurity could get them a job, but no mention of importance of using it in field	It provides a possible career path
Protect my own data or protect myself	Focus on self protection	It could help me keep my computer and information safe
General non-personal answer	Vague platitudes on the importance of cybersecurity	Everything is going digital so it will become increasingly important
No answer or nothing or I don't know		

Fig. 3 shows the proportion of students in each course who gave answers in each theme in Fall 22. For example, 7.4% of CS1 students gave a response with a "Write Secure Software" theme.

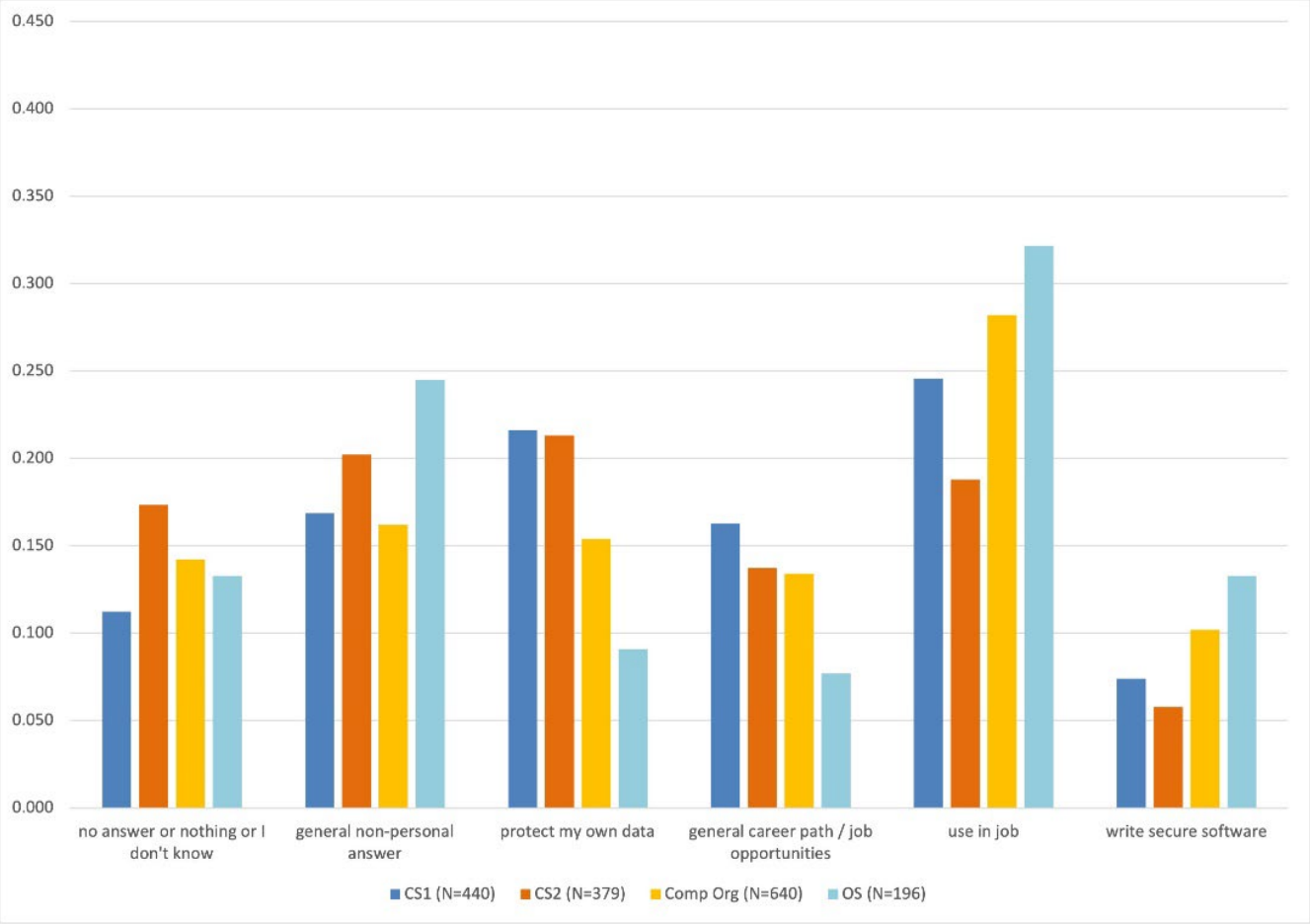

Fig. 3. Fall 2022 "How Is Learning Cybersecurity Relevant to Your Future" Results

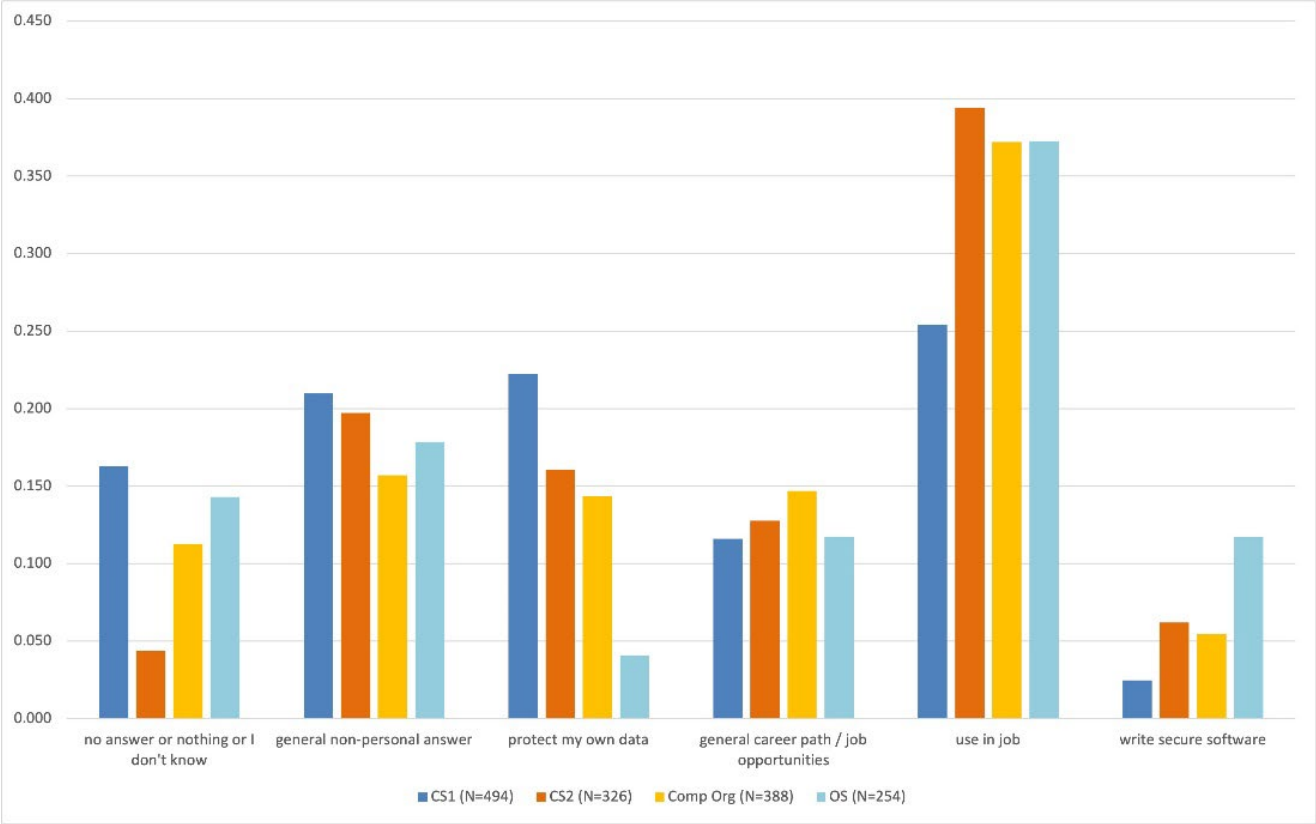

Fig. 4. Spring 2023 "How Is Learning Cybersecurity Relevant to Your Future" Results

Fig. 4 shows the proportion of students in each course who gave answers in each theme in Spring 23.

These data show interesting trends. In Fall 2022 and Spring 2023, "protect myself" was highest for students in CS1, and decreased for each subsequent course. When comparing the proportion of Fall 2022 CS1 students who provided a response with the theme "protect myself" with Fall 2022 CompOrg students who provided a response with this theme, the Z value is -2.30, indicating that CS1 students and CompOrg students are significantly different in their proportion who think that the relevance of studying cybersecurity is about protecting themselves. When comparing Fall 2022 CS1 students with Fall 2022 OS students, Z=-3.27, also indicating that these two groups of students are significantly different in their proportion that mention "protect myself" as the relevant reason to study cybersecurity. These trends are the same in Spring 2023. The difference in proportion for Spring 2023 CS1 students and Spring 2023 CompOrg students has a Z value of -2.62, and for Spring 2023 CS1 students and Spring 2023 OS students is -5.63. Thus, Spring 2023 CS1 students are significantly different than Spring 2023 CompOrg and OS students in their proportion answering "protect myself" as the relevant reason for studying cybersecurity.

Next, "write secure software" is relatively high for students in OS in both Fall 2022 and Spring 2023, and highest for students in CSA in Fall 2022. Students who have been in the curriculum for at least three years are more likely to understand the need for writing secure software, as are students who are in an advanced freshman programming class. The difference in proportion for Fall 2022 CS1 students and Fall 2022 OS students has a Z value of 2.05, indicating that the difference is significant. The Z parameter for the difference in proportion for Spring 2023 CS1 students and Spring 2023 CompOrg students is 2.06, and the Z parameter for the difference between Spring 2023 CS1 and OS students is 4.67. The proportion of CS1 students giving responses in the "write secure software" category is significantly smaller than that of CompOrg and OS students.

Next we look at the differences in the proportions after assignments on input validation were introduced into several courses in our curriculum. In Spring 2023, students in CS2 are likely to have been in CS1 in Fall 2022 and had an assignment where they learned about input validation vulnerabilities. Assignments were also introduced in Software Engineering and CompOrg, so students in OS in Spring of 2023 may have had these assignments. Differences in proportions for OS students Fall 2022 and Spring 2023 do not have any significant differences. CS2 students in Spring 2023 did have some significant differences compared to CS2 students in Fall 2022, however. The Z parameter for CS2 students giving no answer or stating "no" or "nothing" is -4.87 when comparing Fall 2022 CS2 students with Spring 23 CS2 students. A significantly smaller proportion of CS2 students in Spring 2023 gave no answer or said nothing was

relevant, compared to CS2 students in Fall 2022. The Z parameter for CS2 students giving no answer or stating "use in job" is 5.35 when comparing Fall 2022 CS2 students with Spring 2023 CS2 students. A significantly larger proportion of CS2 students in Spring 2023 said cybersecurity was relevant because it could be used in their job, compared to CS2 students in Fall 2022.

VI. DISCUSSION

First we will address RQ1, what do students find interesting or relevant about cybersecurity. Over two semesters, students in core courses at our R1 university were fairly consistent in that the most prevalent subjects mentioned were cryptography, penetration testing, how hacking works, and keeping data safe. This is consistent with English and Maguire [12]. When designing ways to introduce cybersecurity topics to the curriculum, it may be helpful to tie in discussion of these topics. 25% gave no answer at all, or reported that nothing is interesting.

Next, student responses to the relevance of studying cybersecurity included themes of protecting personal data, job opportunities, and writing secure software. About 40-45% of participants thought that learning cybersecurity would help them in their job or write secure code. English and Maguire asked a more specific question and found that 73% thought it would be useful in their career. Our lower numbers could be because English and Maguire surveyed students in their fourth year, while we surveyed students at multiple points in their curriculum. English and Maguire reported 27% did not think it would be useful. Similarly, we found that about 20% gave no answer or answered "nothing". These results do indicate that there is some work to do to convince the 20-25% of students who do not find anything interesting or relevant about cybersecurity.

Next we will address RQ2, how CS students' attitudes differ according to where they are in the curriculum. We found that students in different classes were fairly consistent in what they found interesting about cybersecurity. However, there were interesting patterns among students in the different classes in what they find relevant about studying cybersecurity. Students early in the curriculum have the highest proportion of responses with the theme "protect myself", and the proportion goes down for each subsequent course in the curriculum. Students early in the curriculum have the lowest proportion of responses with the theme "write secure software," and the proportion goes up for subsequent courses. Students early in the curriculum are more likely to focus on the security of their own data, and as they progress, begin to understand that what they create may have security consequences for others. All users of the internet should be concerned about the security of their personal data. It is appropriate for CS students to be concerned about this. However, as future software developers, CS students also need to be concerned with the security of the software that they are creating for others to use. When designing lessons for CS students, it is important to remind them of the need to write software without vulnerabilities. This unique look at how attitudes develop as students progress through the curriculum indicates that students come in with a healthy concern about cybersecurity. However, we could teach them starting at the beginning of the curriculum about their responsibility for learning good cybersecurity practices when writing code.

Finally, we will address RQ3, do CS students' attitudes differ if they have been exposed to cybersecurity learning interventions. Students in several core courses were given assignments on input validation after taking the survey in Fall 2022. Comparing survey results in Fall 2022 to the survey results in Spring 2023 may give insight into the changes in attitude after doing the assignments. In terms of what students find interesting, the only change from Fall 2022 to Spring 2023 is a larger proportion of students responding with "keeping data safe". This may be a result of assignments in some classes, or may be the result of taking the survey a second time. Unfortunately, the number of students giving no answer or stating "nothing" did not change. We hoped that being introduced to cybersecurity assignments (and the survey) would have caused more students to find something interesting about cybersecurity. In terms of the change in what students find relevant about cybersecurity, we did see some changes from Fall 2022 to Spring 2023. CS2 students in Fall 2022 did not have an input validation assignment in their CS1 class, but CS2 students in Spring 2023 did have such an assignment. We found a significantly smaller proportion of CS2 students in Spring 2023 found nothing relevant about studying cybersecurity compared to CS2 students in Fall 2022. We found a significantly larger proportion of CS2 students in Spring 2023 responding that they would use cybersecurity in their job compared to CS2 students in Fall 2022. This is a very promising result, indicating some shift in attitude of CS2 students who learned about input validation. This indicates that we may be able to bring up the fraction of students who understand and appreciate cybersecurity and their role in creating secure software.

VII. LIMITATIONS AND THREATS TO VALIDITY

Participants volunteer to participate in the survey, and thus are self selected. The survey participants are not a true random sample of all students.

Students in later classes in the curriculum are more mature than those in earlier classes. They could have grown in their knowledge and interest in cybersecurity as a result of maturity and experience, rather than as a result of the interventions.

Participants' survey responses are not graded, and they receive extra credit if they answer all the questions. Thus, a reflection response of "nothing" receives the same credit as a detailed response. Willingness to answer a free response question is an imprecise measure of interest in the topic.

For the "interesting" question, about 8% of responses did not match the theme of the cluster in which they were placed, and 5-8% of responses were in a cluster with no common theme. For the "relevant" question, about 16% of responses did not match the theme of the cluster in which they were

placed. The miscategorized responses were not biased toward a particular theme; they were spread evenly across all themes. The proportions are subject to error, but the trends are valid.

VIII. CONCLUSIONS

We found that NLP can be used to find themes in student responses and give valuable insight into what students find interesting and relevant about cybersecurity. We found that students are interested in learning about cryptography and penetration testing, similar to earlier results. We also found that as students progress through the curriculum, more of them realize their unique responsibilities as a CS major and future software developer. When developing lessons and assignments on cybersecurity, it is important to point out the unique responsibilities of coders so that more students make this realization. We found that the proportion of students who saw cybersecurity as relevant to their job duties increased after having lessons and assignments on input validation. In future work we will continue to survey students as they continue through the curriculum and are exposed to multiple lessons in different classes.

REFERENCES

[1] G. Bassett, D. Hylender, P. Langlois, A. Pinto, and S. Widup, "Data breach investigations report," Verizon, Tech. Rep., 2020. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir/

[2] A. Gueye and P. Mell, "A historical and statistical study of the software vulnerability landscape," in *The Seventh International Conference on Advances and Trends in Software Engineering (SOFTENG 2021)*, 2021.

[3] "Mitre common vulnerabilities and exposures," https://cve.mitre.org/cve/data_feeds.html, accessed: 2022-12-16.

[4] ACM, "Curriculum Guidelines for Undergraduate Degree Programs in Computer Science 2013."

[5] X. Yuan, L. Yang, B. Jones, H. Yu, and B.-T. Chu, "Secure Software Engineering Education: Knowledge Area, Curriculum and Resources," *Journal of Cybersecurity Education, Research and Practice*, vol. 2016, no. 1, p. 3, 2016.

[6] A. Farooq, J. Isoaho, S. Virtanen, and J. Isoaho, "Information security awareness in educational institution: An analysis of students' individual factors," *Proceedings - 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2015*, vol. 1, pp. 352–359, 2015, publisher: IEEE ISBN: 9781467379519.

[7] E. B. Kim, "Recommendations for information security awareness training for college students," *Information Management and Computer Security*, vol. 22, no. 1, pp. 115–126, 2014.

[8] A. Farooq, S. Kakakhel, U. Rameez, S. Virtanen, and J. Isoaho, "A taxonomy of perceived information security and privacy threats among IT security students," *2015 10th International Conference for Internet Technology and Secured Transactions, ICITST 2015*, pp. 280–286, 2016, publisher: Infonomics Society ISBN: 9781908320520.

[9] N. Bhatnagar and M. Pry, "Student attitudes, awareness, and perceptions of personal privacy and cybersecurity in the use of social media: An initial study," *Information Systems Education Journal*, vol. 18, no. 1, 2020.

[10] R. Chandarman and B. Van Niekerk, "Students' Cybersecurity Awareness at a Private Tertiary Educational Institution," *The African Journal of Information and Communication*, no. 20, pp. 133–155, 2017.

[11] A. A. Gabra, M. B. Sirat, S. Hajar, and I. B. Dauda, "Cyber security awareness among university students: A case study," *Journal of Critical Reviews*, vol. 7, no. 16, pp. 825–833, 2020.

[12] R. English and J. Maguire, "Exploring student perceptions and expectations of cyber security," in *Computing Education Practice*, 2023, pp. 25–28.

[13] T. Georgescu, "Natural language processing model for automatic analysis of cybersecurity-related documents," *Symmetry*, vol. 12, no. 3, 2020.

[14] "Learning domain-specific word embeddings from sparse author = Roy, A. and Pan, S., note = arXiv: 1709.07470, year = 2020,."

[15] H. Gasmi, J. Laval, and A. Bouras, "Information Extraction of Cybersecurity Concepts: An LSTM Approach," *Applied Sciences*, vol. 9, no. 19, p. 3945, Sep. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/3945

[16] O. De Lin, S. Gottipati, L. S. Ling, and V. Shankararaman, "Mining Informal & Short Student Self-Reflections for Detecting Challenging Topics – A Learning Outcomes Insight Dashboard," in *2021 IEEE Frontiers in Education Conference (FIE)*. Lincoln, NE, USA: IEEE, Oct. 2021, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/9637181/

[17] M. Dorodchi, A. Benedict, and E. Al-Hossami, "Clustering Students' Short Text Reflections: A Software Engineering Course Case Study," p. 7.

[18] "Ap computer science a," https://apstudents.collegeboard.org/courses/ap-computer-science-a, accessed: 2023-7-4.

[19] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "MPNet: Masked and Permuted Pre-training for Language Understanding," p. 11.

[20] "Hugging face: The ai community building the future," https://huggingface.co/, accessed: 2022-12-19.

[21] D. Deng, "DBSCAN Clustering Algorithm Based on Density," in 2020 *7th International Forum on Electrical Engineering and Automation (IFEEA)*. Hefei, China: IEEE, Sep. 2020, pp. 949–953. [Online]. Available: https://ieeexplore.ieee.org/document/9356727

[22] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, Aug. 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3068335

[23] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[24] R. G. Lomax and D. L. Hahs-Vaughn, *An introduction to statistical concepts*. Routledge, 2013.