EPiC Series in Education Science
Volume 5, 2023, Pages 89–103
Proceedings of NEMISA Digital Skills Conference 2023: Scaling Data Skills For Multidisciplinary Impact

Instructors' perception of the competencies required to teach DS in HEI

Nkosikhona Theoren Msweli
University of Pretoria, South Africa.
Nkosikhona.msweli@gmail.com

Abstract

While data science education (DSE) may be a solution for democratising data science in higher education institutions (HEI), challenges on achieving this goal remain. One of the main challenges is the shortage of qualified academic staff members who can deliver multidisciplinary curricula effectively. Teaching data science (DS) is a developing topic that represents a gap in the literature as the education sector embarks on a journey of unearthing the knowledge required to teach DS. This study aimed to gain insight into instructors' perceptions of their skills and competencies in teaching DS. Twenty-six (26) DS instructors were surveyed. The sample included instructors from various disciplines possessing various levels of teaching experience. The study used a 16-construct questionnaire based on the pedagogical content (PCK) framework to understand the scope of DS PCK. The collected data were analysed using IBM SPSS AMOS version 21. Follow-up interviews were conducted to identify competencies needed in DSE and pedagogical beliefs. This qualitative data was analysed thematically. As expected, the study revealed lower female DS instructor participation. The study further showed that the majority of DS instructors have pedagogical and content knowledge of DSE. Business understanding is often taught at the post-graduate and undergraduate levels. Data understanding is taught at the undergraduate level but not as often at the postgraduate level. Short learning programs do not target model evaluation and deployment as a course. The study recommends that curriculum developers consider including the model evaluation and deployment in DS curricula and how these concepts can be taught as part of the curricula. Furthermore, the study recommends the adoption of a DS framework that will guide the development and structuring of DS curricula to ensure standardization and interdisciplinary pedagogies that support content delivery. Research should also be conducted on how industry partnerships can be forged to keep instructors engaged with ongoing DS developments. Lastly, societal interests in data continue to put additional pressure on DSE. Given the field's multidisciplinarity and government's involvement in data usage from various sources, government and industry should play an active role in ensuring the availability, access, and ethical use of data, especially for educational purposes. It is envisaged that this will trigger the need to equip society at large with data skills and policies to govern the process.

1 Introduction

DSE is an emerging and challenging area. It is also a growing topic of interest in Information System (IS) research and education. Several disciplines and complex concepts are blended into DS, and therefore require specific teaching practices to make learning and understanding easier for students. For instance, Machine learning (ML) has many complex and difficult-to-teach algorithms (Sulmont, Patitsas & Cooperstock, 2019) which require DS instructors to empathize with student with diverse backgrounds and expectations (Kross et al., 2022). Irrespective of the complexity of the content or concepts involved, instructors need to ensure that learning outcomes are achieved. This means that DSE instructors need to be agile and open to multiple teaching practices.

DSE programs are becoming more openly available, however, only a few instructors are qualified and knowledgeable enough to teach in this field. This implies that data scientists are generally not properly skilled (Attwood et al., 2019; Yu & Hu, 2019). In fact, the lack of solid background and knowledge in DS among instructors ranks as the biggest limiting factor for integrating DS skills into the curriculum (Emery et al., 2021; Saddiqa et al., 2021). As a result, students enrolled in DS programs are often confronted with challenges emanating from and associated with variable quality of content delivery (Fox & Hackerman 2003; Sunal et al. 2004). It is quite evident that instructors need to find a way to assist students to develop an accurate, adequate, and generally better understanding of DS (Qian et al., 2017). Despite such an obvious gap in DSE, literature that focuses on the design and adoption of strategic approaches for delivering DS programs to a diverse group of students in various domains is scant (Sulmont, Patitsas & Cooperstock, 2019; Twinomurinzi et al., 2022). In addition, research on how DS should be taught or what type of competencies are required from data science academic staff is still lacking. Such research opportunities have the potential to support academia as it struggles to position itself within the data science field (Cao, 2019; Engel, 2017; Mike, 2020) and lead to the creation of new research topics in IS (Cao, 2017) such as how instructors approach DS teaching practices (Lau et al., 2022). Another challenge confronting the field of DSE includes data science tools and techniques that are continuously evolving. Consequently, it is important to examine how instructors tackle this challenge and ultimately unearth instructors' practices as well as their underlying reasoning for teaching DS.

To improve instructors’ knowledge of DS topics, it is important to establish the instructors’ base knowledge, experience, perceptions, and knowledge gap and variation (Saeli et al., 2011; Shulman, 1986). Instructors need a specific type of knowledge to teach DS concepts effectively, and this knowledge is totally different from content knowledge or general pedagogy. Such knowledge is described as PCK (Pedagogy Content Knowledge), that is, a type of knowledge that represents the blending of content (e.g., algorithms, modeling, business scenarios, etc.) and pedagogy (e.g., how to teach algorithms or business cases, etc.). PCK includes the understanding of how instructors will take a specific topic, rearrange it to fit the diverse interests and abilities of learners, and present it for learning purposes (Shulman, 1987).

Understanding how instructors teach DS may also depend on how familiar they are with using DS skills (Emery et al., 2021). Research must address questions such as “what works in DSE” and “what conceptual frameworks guide the practice of DS instructors and enable them to recognise and discuss effective teaching practices?”. Therefore, this study formulated the following research questions:

How does the PCK framework resonates with the competencies of DS lecturers in HEI?

This study aims to gain insight into instructors’ perceptions of their skills and competencies in teaching DS. The PCK framework was adopted to capture some of the essential attributes of knowledge required by facilitators for scholarly integration in their teaching.

The remainder of the paper is structured as follows: following this introduction, the study takes a brief look at the study’s theoretical framework. Thereafter, we give an account of the method used to collect the empirical data used in the study. After presenting the results and discussing the contribution and limitations of the study, the paper concludes by addressing the implications of the results for DSE and suggesting an agenda for further research.

2 Literature Review

2.1 DS Instructional Programs

At the postgraduate level, DS programs have been proliferating across the globe (Hosack and Sagers, 2015; Raj et al., 2019; Li, Milonas & Zhang, 2021). On the other hand, undergraduate DS programs are still being investigated (Zhang, Huang & Wang, 2017; Mikalef et al., 2018; Çetinkaya-Rundel & Ellison, 2021; Davenport & Malone, 2021). DS short-learning programs are often well designed and commoditized to mainly address DS technical skills such as predictions, data analytics, ML, and statistical programming. However, the integration of these technical concepts within a full DS course still needs to be explored (Qiang et al., 2019; Silva et al., 2014). In particular, the development of teaching guidelines for DSE and training has not been adequately researched.

As attested by Demchenko et al. (2019), DSE must reflect multi-disciplinary knowledge and competencies to afford data scientists insights into other domains. DSE should further afford skills and competencies to work with various forms of data and interpret the analytical results, especially for those who lack DS literacy (Dichev & Dicheva, 2017). Therefore, it is important to establish instructors' confidence in teaching DS (Mike, 2020), or how best to prepare instructors to teach DS in various domains (Emery et al., 2021). For instance, data scientists in biomedicine need to be trained in computer science, statistics and mathematics, and biomedicine (Garmire et al., 2017; Hassan & Liu, 2020). Stephenson et al. (2018) alluded to the fact that offering DS skills in computer science courses only could lead to the under-representation of other disciplines. Research on how to address teaching practices in DSE are necessary. For instance, Emery et al. (2021) investigated ways of preparing instructors to teach DS in undergraduate biology and environmental science courses.

2.2 Teaching DS

Fayyad and Hamutcu (2021) raised an important question "how do we teach data scientists while there is so much debate on who they are?". There appears to be a growing concern among those who provide training and employment to data scientists. Apart from unclear roles of data scientists, teaching DS faces other challenges, including teaching multidisciplinary content, the misconception of DS concepts (Jafar, Babb & Abdullat, 2016), student diversity, and student cognitive skills (Sulmont, Patitsas & Cooperstock, 2019; Donoghue et al., 2021). These challenges may lead to low student registrations and throughput in DS programs. Instructors need to know their students and their characteristics to apply appropriate pedagogy (Gudmundsdottir & Shulman, 1987). Sentance and Csizmadia (2017) found that various pedagogies can improve students' ability to solve a problem. However, instructors need support in terms of professional development on how to work with different pedagogies in a multidisciplinary setting to effectively teach DS concepts (Emery et al., 2021; Lau et al., 2022). Table 1 provides examples of pegagogies used for specific concepts.

Data science concept	***Teaching pedagogy/strategy***	***Source***
Data preparation, Visualisation	Project-based learning	Saltz and Heckman (2015)

Machine learning, modeling	Similes, gamification, storytelling	Song and Zhu (2016; Garcia-Algarra (2020)
Model deployment	Story-telling Experiential learning	Jaggia *et al.* (2020) Anslow *et al.* (2016)

Table 1:Pedagogies in DSE

Effective teaching occurs when learning and understanding are achieved (Blair et al., 2021). To determine whether student learning has been enhanced, teachers must evaluate their teaching practices. DS instructors interested in assessing their teaching practices can apply various frameworks to inform their questions and teaching strategies (Kim, Ismay & Chunn, 2018; Hassan & Liu, 2019). It is worth noting that the conceptual framework that is applicable and useful to higher education; the focus has instead been on secondary schools (Saeli et al., 2011; Başaran, 2020; Taopan, Drajati & Sumardi, 2020).

While some DS concepts appear to have existed in various domains, strategies for their content delivery are lacking (Sulmont, Patitsas & Cooperstock, 2019; Dill-McFarland et al., 2021; Lau et al., 2022). For instance, data frames that were originally designed by statisticians for exploratory data analysis, are now viewed by data scientists as data sets. Instructors can discuss how each discipline or any domain for that matter views and uses a particular concept (Jafar, Babb & Abdullat, 2016; Lau et al., 2022). Technical concepts, especially the ones listed in Table 2, have been put forward as key competencies of DS, and instructors are expected to demonstrate these competencies to teach DS.

Table 2: DS competencies

Data science competency	Source
Data visualisation	Shirani (2016)
Apache Hadoop and programming languages	Demchenko (2019; Price & Ramaswamy (2019; Yadav & Debello (2019)
Machine learning	Garcia-Algarra (2020)
Big Data and ethics	Saltz, Dewar & Heckman (2018); Mike (2020)

2.3 The PCK theoretical framework

The concept of PCK was introduced by Shulman (1986), after pointing out a lack of research that targets the course content taught to students. Shulman, (1986) defined pedagogical content knowledge as teachers' interpretations and transformations of subject-matter knowledge in the context of facilitating student learning. As shown in Figure 1, integrating content knowledge (CK) and pedagogy knowledge (PK) enables an understanding of how particular topics are presented to students with different backgrounds. Instructors should be able to transform the knowledge to be taught to the students in a way that is easily understood.

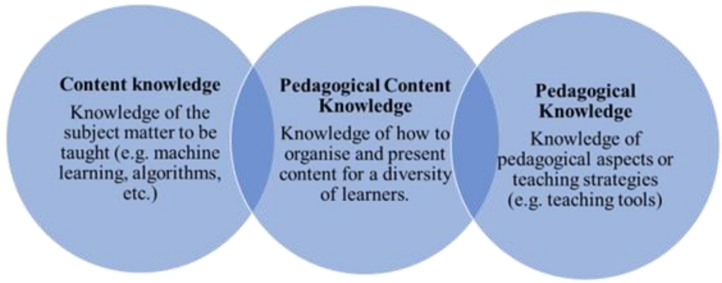

Figure 1: An illustration of pedagogical content knowledge *Source: Shulman (1986)*

PCK outlines the instructor's domain knowledge, pedagogical knowledge, and knowledge of the environment in which the subject is being taught. The framework is useful when differentiating teaching- and non-teaching specialists, for instance, a data science instructor from a data scientist. The argument lies in the capacity of the instructor to transform the CK into forms that are pedagogically powerful and yet adaptive to student backgrounds and abilities (Shulman, 1987). The key elements of pedagogical content knowledge proposed by Shulman, (1987) are as follows:

1. **Content Knowledge** - Knowledge of representations of subject matter – it is about the actual subject matter that is to be taught. Instructors must know and understand the subjects that they teach, including knowledge of methods, tools, concepts, theories, and techniques within a given field. For example, in DSE, instructors know that students must learn ML, algorithms, analytics, data visualization, and so on. Therefore, CK is required for knowledge and understanding of how these concepts or areas can be and are being taught, and the advantages and disadvantages of each teaching practice.

2. **Student Knowledge** – Understanding of students on the subject and the learning and teaching implications that are associated with the specific subject matter. For example, students may confuse data mining and data wrangling as the same concept. Alternatively, students assume business/domain knowledge is not so important.

3. **General pedagogical knowledge (Pedagogical Knowledge)** – Understanding of the practices or methods of teaching and learning and how this understanding encompasses among other things, overall educational purposes, values, and aims.

   Additional Elements:

4. **Curriculum Knowledge** – Knowledge of what should be taught to a particular group of students. Instructors need to know students' learning potential, syllabuses, and program planning documents and how assessments will be conducted.

5. **Pedagogical content knowledge** – Instructors' understanding of how to teach the subject matter, including the use of examples and illustrations to make a particular topic understandable across all students.

6. Knowledge of students and their characteristics and how these may affect their learning,

7. Knowledge of educational contexts, the political, social, and religious workings of groups or of the classroom in which the teaching takes place.

This study adopted PCK to understand the instructor's perceptions of their knowledge of teaching data science. Prior PCK research into computer science education examined topics such as science (Fraser, 2016), programming (Qian et al., 2017; Rahimi et al., 2018; Saeli et al., 2011), and design of digital artifacts (Rahimi, Barendsen, & Henze, 2016). However, there is little scientific understanding and reporting of teachers' PCK for teaching algorithms, except for Sulmont, Patitsas and Cooperstock (2019) who investigated the difficulties of teaching ML to non-STEM (non-science, technology, engineering, and mathematics) students.

3 Methodology

The study focused on instructors' competence to teach DS by using a 16-construct questionnaire that was based on the PCK framework as a way to understand the scope of what DS PCK could be. Data was collected via an online questionnaire. Quantitative data were analyzed using simple descriptive statistics.

3.1 Participants demographics

A total of 26 instructors participated in the study. These participants were selected using purposeful sampling method. Table 3 shows the demographical information of the study participants.

Category	Sub-category	Frequency	Percentage
Sex	Male	21	80.8
	Female	5	19.2
Highest Qualification	Master's Degree	8	30.8
	Honours Degree	7	26.9
	Bachelor's Degree	10	38.5
	Others	1	3.8
Years of Experience	Less than 1 year	4	15.4
	1 to 5 years of Experience	6	23.1
	+5 years of Experience	4	15.4
	+10 years of Experience	12	46.2
Age	18 to 35 years	4	16.7
	36 to 55 years	20	83.3
Level of data science qualification teaching at	Short Learning Programme (e.g., MOOCs, Badges, micro-credentials)	8	30.8
	Undergraduate level (e.g., Bachelor's Degree, Diploma, Higher Certificate)	11	42.3
	Postgraduate level (Honours, Masters, Doctorate)	7	269

Table 3: Demographics of study participants

4 Results and findings

This section provides details on how the data was captured, described, analysed, and interpreted systematically.

4.1 Descriptive statistics

Central tendency measures

Central tendency measures were conducted to assess how centred the distribution of the constructs involved in the study is. A five-point Likert scale where the value 1 corresponds to "Strongly disagree" and the value 5 corresponds to "Strongly agree" was applied to measure the following constructs: Content Knowledge (CK), Pedagogical Knowledge (PK), and Pedagogical Content Knowledge (PCK). Table 4 summarises the responses of the participants with only high frequency in each category being reported.

Category	Questions	measure	Frequency	Percentage
Content Knowledge	I am familiar with data science tools, processes, and technique	To a great extent	11	42.3
	I do understand various concepts of data science	To a great extent	11	42.3
	I know what data science students should be taught in terms of content	To a great extent	11	42.3
	I have knowledge and understanding of data science and what it entails	To a great extent	12	46.2
	I am familiar with the data science curriculum and syllabus	To a great extent	10	38.5
	I can create materials that map to a specific level of proficiency among students in teaching data science	To a great extent	10	38.5
Pedagogical Knowledge	I know of the different processes and practices of teaching e.g. establishing learning objectives	To a large extent	13	50
	I know how to organize a classroom and manage students during instruction	To a great extent	15	57.7
	I can differentiate between various instructional strategies (teaching pedagogies)	To a large extent	12	46.2
	I can use various teaching pedagogies	To a great extent	12	46.2
	I know how to align learning outcomes and assessment opportunities with the teaching pedagogy	To a great extent	14	53.8
	I know how to teach in a multidisciplinary setting	To a large extent	15	57.7

	I can identify different strategies for evaluating student understanding	To a large extent	12	46.2
Pedagogical Content Knowledge	I know how to teach data science topics/concepts using appropriate teaching pedagogies	To a great extent	13	50
	I know how to pair teaching pedagogy with data science concepts when preparing and delivering the content	To a large extent	12	46.2
	I know the teaching pedagogies that are appropriate for data science education	To a large extent	14	53.8

Table 4: Summary of descriptive statistics on PCK

- ***Content Knowledge***

Instructors were asked about their content knowledge of DS. Six questions were posed to participants to establish their level of knowledge. The results indicate that, to a great extent, 46,2% of instructors have the required knowledge and understanding of DS. In the same vein, 42,3% of the instructors were found to be familiar with DS tools, processes, and techniques; understand various concepts of DS; and know what students should be taught in terms of content. The overall mean (4,12) indicates that most instructors have, to a large extent, the requisite content knowledge.

- ***Pedagogical Knowledge***

Instructors were asked about their pedagogical knowledge of DS. Seven questions were asked to establish their level of knowledge. The results indicate that the majority (57,7%) of instructors know to a large extent how to teach in a multidisciplinary setting. Following in the same pattern, 57,7% of the instructors, to a great extent, know how to organize a classroom and manage students during instruction. Furthermore, 53,8% of the instructors know, to a great extent, how to align learning outcomes and assessment opportunities with the teaching pedagogy. Whereas 50% of the instructors were found to be, to a large extent, knowledgeable of the different processes and practices of teaching e.g., establishing learning objectives,; 50% of the instructors could, to a great extent, identify different strategies for evaluating student understanding.

The overall mean results (4,35) indicate that most instructors, to a large extent, have Pedagogical Knowledge.

- ***Pedagogical Content Knowledge***

Instructors were questioned about their pedagogical knowledge of DS, and 3 questions were posed to establish the level of their knowledge. Based on the results generated from, Results of this part of the study revealed that slightly over half (53,8%) of the instructors, to a large extent, know the teaching pedagogies appropriate for DSE. Analogously, 50% of the instructors, to a great extent, know how to teach DS topics/concepts using appropriate teaching pedagogies. Lastly, only 46,2% of the instructors (to a large extent) know how to pair teaching pedagogy with data science concepts when preparing and delivering the content. The mean score (4,13) suggests that most of the instructors, to a large extent, have pedagogical content knowledge.

Cross tabulations

Cross-tabulation enables quantitative analysis of the data to understand the relationship or correlation between multiple variables. Several relationships were studied and the results obtained are discussed below:

- The relationship between business understanding and the level of DS qualification they teach - the results revealed that 80% of instructors that teach at the postgraduate level often teach business requirement (business understanding). In comparison, 50% of instructors that teach at the undergraduate level always teach business understanding.
- The relationship between data understanding and the level of DS qualification they teach - The results indicate that 38.5% of the instructors that teach at the postgraduate level often teach data understanding. In contrast, a substantial majority of the instructors (83.3%) that teach at the undergraduate level always teach data understanding.
- The relationship between data preparation and the level of DS qualification they teach - the results indicate that most (63.6%) of the instructors that teach at the undergraduate level always teach data preparation.
- The relationship between data modeling and the level of DS qualification they teach - the results indicate that less than half (45.5%) of the instructors that teach at the undergraduate level often teach data modeling.
- The relationship between model evaluation and the level of DS qualification they teach - the results show that half of the instructors that teach short learning programmes rarely teach model evaluation. In comparison, a mere 40% of the instructors that teach at the undergraduate level often teach model evaluation.
- The relationship between deployment and the level of DS qualification they teach - the results indicate that 80% of instructors that teach DS at the undergraduate always teach how models are deployed.

5 Discussion

DSE demands an interdisciplinary curriculum (Twinomurinzi et al., 2022), and instructors are compelled to employ multiple pedagogies to deliver this curriculum (Asamoah, Doran & Schiller, 2020). When applying PCK, it is envisioned that instructors incorporate their interdisciplinary pedagogical knowledge into teaching data science. It is expected that instructors with over 10 years of teaching experience have the requisite experience to use different pedagogies to teach the curricula. However, their experience may not pertain to teaching data science considering that it is fairly new and emerging discipline. It should be borne in mind that data science instructors are responsible for researching, preparing, conducting, and reviewing educational programs. That being the case, they are also responsible for developing new skills for data scientists. Essentially, instructors may need to enhance their knowledge to fit current trends and familiarize themselves with the content and its application in the real world. This implies that HEIs need resources to capacitate and develop instructors as new trends emerge. For instance, instructors who may not be acquainted with Auto ML or have expertise in Hadoop and Spark could form part of a continuous development or life-long learning program.

Reasons behind gender-based differences in the adoption and use of technology continue to be a challenge that is not addressed by research. Work on the dominance of technology used by males as compared to females is abound (Dichev & Dicheva, 2017). Shahbaz *et al.* (2020) reported that, when compared with females, males feel data analytics is powerful and more useful. It is evident from the low participation rate of female DS instructors in this study that gender gaps still persist in the field of

technology, especially DSE. Partnerships with different communities need to be explored to improve the
under-representation of women (Gundlach & Ward, 2021). In education, PCK is necessary to determine gender factors impacting the adoption and use of technology (Saeli et al., 2011).

Model evaluation is not often taught as part of the curriculum. While this step is often overlooked, its significance has triggered the need to standardize it (Baier, Jöhren & Seebacher, 2019). Castellanos *et al.* (2019) have noted that while the industry has shown more interest in data science models, the deployment rate of these models is still very low. One of the reasons could be that the models are not evaluated to establish whether they satisfy all business objectives to qualify for deployment. There is also a possibility that the deployment phase and skills applied are not DS-focused (Ackermann *et al.*, 2018; Baier, Jöhren & Seebacher, 2019). This accentuates the importance of a DS project framework to guide the development of DS programs. Furthermore, such an approach will contribute towards addressing technical and non-technical challenges that are often experienced during the deployment stage (Baier, Jöhren & Seebacher, 2019). The results reported herein revealed the gaps in the inclusion of model deployment in DS teaching and learning. Previous work has also identified these factors (Davenport & Malone, 2021). While it is important to build a working model, it is also important to determine how the industry receives these models and how they are deployed to improve their adoption (Ackermann et al., 2018). Previous work indicates that it is difficult to teach the deployment of models in an educational setting (Jaggia et al., 2020). This could imply inexperience or a lack of skills in DS infrastructure (Castellanos et al., 2019).

While computer science undergraduate programs are the popular preference for DS (Mike, 2020), incorporating DS into domain programs can be of immense benefit (Castellanos *et al.*, 2019; Davenport & Malone, 2021). It is important to understand that the nature of DS has distinct needs and significance depending on the organisation or domain. The same can be said of the way DS is taught to science and non-science students. Such an understanding can be achieved through pedagogical advancements which provide new ways to teach DS concepts and thus build a workforce that is industry relevant. In addition, DS offerings at postgraduate level has the potential to elevate the skills levels of students when they are busy with their studies (Hosack & Sagers, 2015). Having knowledgeable instructors has the ability to support the attainment of advanced analytical skills.

Business understanding and data understanding appeared to be more common in undergraduate programs. These two components are crucial in any DS project. One of the important data challenges is that it moves very fast in different cycles thus leading to data skills being outdated rapidly. It is rather necessary to keep up with new trends and establish a relationship with leading industries for joined initiatives on DSE (Mikalef et al., 2018). PCK is considered powerful in influencing the pedagogical thinking that is necessary for DSE. To illustrate this point, instructors' unfamiliarity with data could be alleviated through continuous professional development in a form of short courses (such as micro-credentials or MOOCs) and workshops (Saddiqa et al., 2021). In a fast-paced industry and ever-changing technologies, micro-credentials offer a better flexible solution for skilling individuals (Msweli, Twinomurinzi & Ismail, 2022).

Therefore, academic training, related industry experience, licensure, prior training, or lecturing experience are required. Knowledge of statistics, programming, data visualization, big data, and building models (machine learning) is also required.

6 Conclusion, Implications, Limitations, and Areas for further research

PCK for DS instructors is important since DS programs are becoming easily available and are thus accepting students from various backgrounds. Teaching in this field comes with opportunities and challenges. The purpose of this study was to investigate the extent to which the PCK framework resonated with the competence of DS lecturers and how it influences the teaching practices in DSE. The study examined how the PCK framework can be adopted to assess the instructor's knowledge of teaching DS. Therefore, this paper contributes to the body of knowledge by understanding the confidence and competence of those that are teaching or aspire to teach DS. This contribution presents a gap in DS offerings where there is no guidance on how and at what level the specific DS content should be presented, and how it should be presented. The study immediately found fewer female data science instructors. This prompt the need to study gender differences to find the moderating factors behind the low representation of females in data science educational contexts.

A need still exists to clearly define the roles of data scientists so that they can be trained accordingly. For example, their involvement in model evaluation and deployment is not clear. Research is needed in this regard to establish the data scientist's involvement in data science projects. This study has revealed that PCK is useful to instructors when addressing the following knowledge questions: What are the reasons behind teaching a specific DS program?; What DS concepts should be taught by instructors?; What challenges or misconceptions do students encounter within these concepts; and, How should these concepts be taught? Furthermore, it is useful to understand the target audience of the intended course. The overall data science field is unstable and advancing rapidly. Therefore, the findings and recommendations of this present study include emphasizing the importance of continued learning. Such an approach will enable instructors to acquire new data science knowledge and competencies that currently do not exist, thus making it easier to disseminate the purported new knowledge during lesson delivery. This continued learning can be undertaken or achieved through training and workshops, research, or collaborative partnerships with industry. Faculty heads may need to invest resources to improve the quality of teaching and the confidence of data science instructors. This will further require the cooperation of both HEI management and instructors to ensure effective teaching. This research advocates for the use of the PCK framework to improve the teaching and learning of multidisciplinary data science curricula.

LIMITATIONS AND AREAS FOR FUTURE WORK

The study only considered PCK without exploring other factors that might influence the knowledge of teaching DS. The PCK is content general, it does not apply to any specific subject. Therefore, where there are challenges or complex concepts, PCK might not be relevant. The other limitation is that the sample size used in this study was too small and based on a single under-developed region. Future studies can increase the sample size and focus on other regions that are, for example, well off.
Further work needs to consider various factors and how they affect knowledge.

The study further suggests the following for future research:

- Alignment of DS programs with CRISP-DM or the adoption of other similar frameworks.
- Investigate ways/methods of teaching evaluation and deployment as part of the DS curriculum.
- Investigate which pedagogies work better with DS concepts.
- Enquiry on importance of industry knowledge and experience among DS instructors.
- Gender inequality in a DS discipline.

This research recommends the use of the PCK model in the educational field to plan, organize and carry out DSE activities.

ACKNOWLEDGEMENTS
None.

References

Ackermann, K., Walsh, J., De Unánue, A., Naveed, H., Rivera, A. N., Lee, S.-J., Bennett, J., Defoe, M., Cody, C., Haynes, L., & Ghani, R. (2018). Deploying Machine Learning Models for Public Policy: A Framework *. *KDD*, 15–22. https://doi.org/10.1145/3219819.3219911

Anslow, C., Brosz, J., Maurer, F., & Boyes, M. (2016). Datathons: An Experience Report of Data Hackathons for Data Science Education. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 615–620. https://doi.org/10.1145/2839509.2844568

Asamoah, D. A., Doran, D., & Schiller, S. (2020). Interdisciplinarity in Data Science Pedagogy: A Foundational Design. *Journal of Computer Information Systems*, *60*(4), 370–377. https://doi.org/10.1080/08874417.2018.1496803

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., & Schneider, M. V. (2019). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*, *20*(2), 398–404. https://doi.org/10.1093/bib/bbx100

Baier, L., Jöhren, F., & Seebacher, S. (2019). Challenges in the Deployment and Operation of Machine Learning in Practise. *Twenty-Seventh European Conference on Information Systems (ECIS2019)*. https://www.researchgate.net/publication/332996647

Başaran, B. (2020). nvestigating science and mathematics teacher candidate's perceptions of TPACK-21 based on 21st century skills. *Elementary Education Online*, *19*(4), 2212–2226. https://doi.org/10.17051/ilkonline.2020.763851

Blair, J. R. S., Jones, L., Leidig, P., Murray, S., Raj, R. K., & Romanowski, C. J. (2021). Establishing ABET Accreditation Criteria for Data Science. *SIGCSE 2021 - Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 535–540. https://doi.org/10.1145/3408877.3432445

Cao, L. (2019). Data Science: Profession and Education. *IEEE Intelligent Systems*, *34*(5), 35–44. https://doi.org/10.1109/MIS.2019.2936705

Castellanos, C., Pérez, B., Varela, C. A., Del Pilar Villamil, M., & Correal, D. (2019). A Survey on Big Data Analytics Solutions Deployment. *European Conference on Software Architecture*, 195–210.

Çetinkaya-Rundel, M., & Ellison, V. (2021). A Fresh Look at Introductory Data Science. *Journal of Stactistics and Data Science Education*, *29*(1), 16–26. https://doi.org/10.1080/10691898.2020.1804497

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. *Harvard Data Science Review*, *3*(3). https://doi.org/10.1162/99608f92.90814c32

Demchenko, Y. (2019). Big data platforms and tools for data analytics in the data science engineering curriculum. *ACM International Conference Proceeding Series*, 60–64. https://doi.org/10.1145/3358505.3358512

Dichev, C., & Dicheva, D. (2017). Towards Data Science Literacy. *Procedia Computer Science 108C*, 2151–2160.

Dill-McFarland, K. A., Konig, S. G., Mazel, F., Oliver, D. C., McEwen, L. M., Hong, K. Y., & Hallam, S. J. (2021). An integrated, modular approach to data science education in microbiology. *PLoS Computational Biology*, *17*(2). https://doi.org/10.1371/JOURNAL.PCBI.1008661

Donoghue, T., Voytek, B., … S. E. S. and D. S., & 2021, U. (2021). Teaching creative and practical data science at scale. *Taylor & Francis*, *29*(S1), 27–39. https://doi.org/10.1080/10691898.2020.1860725

Emery, N., Supp, S. R., Kerkhoff, A. J., Farrell, K. J., Bledsoe, E. K., Mccall, A. C., & Aiello, M. (2021). Training Data: How can we best prepare instructors to teach data science in undergraduate biology and environmental science courses? *BioRxiv*. https://doi.org/10.1101/2021.01.25.428169

Engel, J. (2017). Statistical Literacy for Active Citizenship: a Call for Data Science Education. *Statistics Education Research Journal*, *16*(1), 44–49. http://iase-web.org/Publications.php?p=SERJ

Fayyad, U., & Hamutcu, H. (2021). How Can We Train Data Scientists When We Can't Agree on Who They Are? *Harvard Data Science Review*, *3*, 1–7. https://doi.org/10.1162/99608f92.0136867f

Fraser, S. P. (2016). Pedagogical Content Knowledge (PCK): Exploring its Usefulness for Science Lecturers in Higher Education. *Res Sci Educ*, *46*, 141–161. https://doi.org/10.1007/s11165-014-9459-1

Garcia-Algarra, J. (2020). Introductory Machine Learning for Non STEM Students. *Proceedings of the European Conference on Machine Learning*.

Garmire, L. X., Gliske, S., Nguyen, Q. C., Chen, J. H., Nemati, S., Van Horn, J. D., Moore, J. H., Shreffler, C., & Dunn, M. (2017). The Training of next Generation Data Scientists in Biomedicine. *Pacific Symposium on Biocomputing 2017*. www.worldscientific.com

Gudmundsdottir, S., & Shulman, L. (1987). Pedagogical Content Knowledge in Social Studies. *Scandinavian Journal of Educationl Research*, *31*(2), 59–70. https://doi.org/10.1080/0031383870310201

Gundlach, E., & Ward, M. D. (2021). The Data Mine: Enabling Data Science Across the Curriculum. *Journal of Statistics and Data Science Education*, *29*(S1), S74–S82. https://doi.org/10.1080/10691898.2020.1848484

Hassan, I. B., & Liu, J. (2019). Embedding data science into computer science education. *IEEE International Conference on Electro Information Technology*, *2019-May*, 367–372. https://doi.org/10.1109/EIT.2019.8833753

Hassan, I. B., & Liu, J. (2020). A Comparative Study of the Academic Programs between Informatics/BioInformatics and Data Science in the U.S. *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, 165–171. https://doi.org/10.1109/COMPSAC48688.2020.00030

Hosack, B., & Sagers, G. (2015). Applied Doctorate in IT: A Case for Designing Data Science Graduate Programs. *Journal of the Midwest Association for Information Systems (JMWAIS)*, *1*(1), 1. http://aisel.aisnet.org/jmwais/vol1/iss1/6

Jafar, M. J., Babb, J., & Abdullat, A. (2016). Emergence of Data Analytics in the Information Systems Curriculum. *Proceedings of the EDSIG Conference*. http://iscap.info

Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2020). Applying the CRISP-DM Framework for Teaching Business Analytics. *Decision Sciences Journal of Innovative Education*, *18*(4), 612–634. https://doi.org/10.1111/dsji.12222

Kim, A. Y., Ismay, C., & Chunn, J. (2018). The fivethirtyeight R package: "Tame Data" Principles for Introductory Statistics and Data Science Courses Chester Ismay Recommended Citation. *Technology Innovations in Statistics Education*, *11*(1). https://scholarworks.smith.edu/mth_facpubs/47

Kross, S., Peng, R. D., Caffo, B. S., Gooding, I., & Leek, J. T. (2022). The Democratization of Data Science Education. *The American Statistician*, *74*(1), 1–7. https://doi.org/10.1080/00031305.2019.1668849

Lau, S., Nolan, D., Gonzalez, J., & Guo, P. J. (2022). How Computer Science and Statistics Instructors Approach Data Science Pedagogy Differently: Three Case Studies. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, *1*, 29–35. https://doi.org/10.1145/3478431.3499384

Li, D., Milonas, E., & Zhang, Q. (2021, August 20). Content Analysis of Data Science Graduate Programs in the U.S. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2--36841

Mikalef, P., Giannakos, M. N., Pappas, I. O., & Krogstie, J. (2018). The human side of big data: Understanding the skills of the data scientist in education and industry. *IEEE Global Engineering Education Conference, EDUCON, 2018-April*, 503–512. https://doi.org/10.1109/EDUCON.2018.8363273

Mike, K. (2020). Data Science Education: Curriculum and pedagogy. *ICER 2020 - Proceedings of the 2020 ACM Conference on International Computing Education Research*, 324–325. https://doi.org/10.1145/3372782.3407110

Msweli, N. T., Twinomurinzi, H., & Ismail, M. (2022). The International Case for Micro-Credentials for Life-Wide And Life-Long Learning: A Systematic Literature Review. *Interdisciplinary Journal of Information, Knowledge, and Management*, *17*, 151–190. https://doi.org/10.28945/4954

Price, R., & Ramaswamy, L. (2019). Challenges and approaches to teaching data science technologies in an information technology program with non-traditional students. *Proceedings - 2019 IEEE 5th International Conference on Collaboration and Internet Computing, CIC 2019*, *Cic*, 49–56. https://doi.org/10.1109/CIC48465.2019.00015

Qian, Y., Lehman, J., Qian, Y., & Lehman, J. (2017). Students' Misconceptions and Other Difficulties in Introductory Pro-gramming: A Literature Review. *ACM Transactions on Computing Education*, *18*(1). https://doi.org/10.1145/3077618

Qiang, Z., Dai, F., Lin, H., & Dong, Y. (2019). Research on the Course System of Data Science and Engineering Major. *2019 IEEE International Conference on Computer Science and Educational Informatization, CSEI 2019*, 90–93. https://doi.org/10.1109/CSEI47661.2019.8938944

Rahimi, E., Henze, I., Hermans, F., & Barendsen, E. (2018). Investigating the pedagogical content knowledge of teachers attending a MOOC on scratch programming. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11169 LNCS*, 180–193. https://doi.org/10.1007/978-3-030-02750-6_14/TABLES/2

Raj, R. K., Parrish, A., Impagliazzo, J., Romanowski, C. J., Aly Ahmed, S., Bennett, C. C., Davis, K. C., McGettrick, A., Susana Mendes Pereira, T., & Sundin, L. (2019). Data Science Education: Global Perspectives and Convergence. *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*, *19*. https://doi.org/10.1145/3304221.3325533

Saddiqa, M., Magnussen, R., Larsen, B., & Pedersen, J. M. (2021). Open Data Interface (ODI) for secondary school education. *Computers and Education*, *174*, 104294. https://doi.org/10.1016/J.COMPEDU.2021.104294

Saeli, M., Perrenet, J., Jochems, W. M. G., & Zwaneveld, B. (2011). Teaching Programming in Secondary School: A Pedagogical Content Knowledge Perspective. *Informatics in Education*, *10*(1), 73–88.

Saltz, J., & Heckman, R. (2015). Big Data science education: A case study of a project-focused introductory course. *Themes in Science & Technology Education*, *8*(2), 85–94. https://www.learntechlib.org/p/171521/

Saltz, J. S., Dewar, N. I., & Heckman, R. (2018). Key Concepts for a Data Science Ethics Curriculum. *ACM Reference*. https://doi.org/10.1145/3159450.3159483

Sentance, S., & Csizmadia, A. (2017). Computing in the curriculum: Challenges and strategies from a teacher's perspective. *Education and Information Technologies*, *22*, 469–495. https://doi.org/10.1007/s10639-016-9482-0

Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., & Arshad, M. R. (2020). Moderating Effects of Gender and Resistance to Change on the Adoption of Big Data Analytics in Healthcare. *Complexity*, 1–13. https://doi.org/10.1155/2020/2173765

Shirani, A. (2016). Identifying Data Science and Analytics Competencies Based on Industry Demand. *Issues in Information Systems*, *17*(IV), 137–144. https://doi.org/10.48009/4_iis_2016_137-144

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *14*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23.

Silva, Y. N., Dietrich, S. W., Reed, J. M., & Tsosie, L. M. (2014). Integrating Big Data into the computing curricula. *SIGCSE 2014 - Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, 139–144. https://doi.org/10.1145/2538862.2538877

Song, I. Y., & Zhu, Y. (2016). Big data and data science: what should we teach? *Expert Systems*, *33*(4), 364–373. https://doi.org/10.1111/exsy.12130

Sulmont, E., Patitsas, E., & Cooperstock, J. R. (2019). What is hard about teaching machine learning to non-majors? Insights from classifying instructors' learning goals. *ACM Transactions on Computing Education*, *19*(4), 1–16. https://doi.org/10.1145/3336124

Taopan, L. L., Drajati, N. A., & Sumardi. (2020). TPACK Framework: Challenges and Opportunities in EFL Classrooms. *Research and Innovation in Language Learning*, *3*(1), 1–22. https://doi.org/10.33603/RILL.V3I1.2763

Twinomurinzi, H., Mhlongo, S., Bwalya, K. J., Bokaba, T., & Mbeya, S. (2022). Multidisciplinarity in Data Science Curricula. *African Conference on Information Systems and Technology*.

Yadav, N., & Debello, J. E. (2019). Recommended Practices for Python Pedagogy in Graduate Data Science Courses. *Proceedings - Frontiers in Education Conference, FIE*, *2019-Octob*. https://doi.org/10.1109/FIE43999.2019.9028449

Yu, B., & Hu, X. (2019). Toward Training and Assessing Reproducible Data Analysis in Data Science Education. *Creative Commons Attribution 4.0 International (CC BY 4.0)*, *1*, 381–392. https://doi.org/10.1162/dint_a_00053

Zhang, X., Huang, X., & Wang, F. (2017). The construction of undergraduate data mining course in the big data age. *ICCSE 2017 - 12th International Conference on Computer Science and Education*, *Iccse*, 651–654. https://doi.org/10.1109/ICCSE.2017.8085573

Appendices

https://bit.ly/DSEInstructorPCK