Knowledge and Skill Sets for Big Data Profession: Assessing Student's Quality using Exploratory Factor Analysis

Sarah Yusoff [1*], Nur Hidayah Md Noh[2], Norulhidayah Isa[3], Siti Musliha Nor-Al-din[4]
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA Cawangan Terengganu,
Kampus Kuala Terengganu, Terengganu, Malaysia
*sarahyusoff@uitm.edu.my[1], norhidayah0378@uitm.edu.my[2], norul955@uitm.edu.my[3], sitim907@uitm.edu.my[4]

2022 International Visualization, Informatics and Technology Conference (IVIT) | 978-1-6654-8648-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/IVIT55443.2022.10033399

***Abstract*— Recently, several higher education institutions in Malaysia announced discontinuing some courses to ensure employability post-graduation. Finding a job that fits their qualifications is a hurdle that graduates frequently face. The International Labor Organization states that when the education and training system does not deliver the skills the labour market needs, there is a mismatch between skills and jobs. This paper presents research on big data analytics knowledge and skills acquired by students throughout their studies. A sample of 185 UiTM students from various campuses participated. These students were among those who had formally taken big data courses during their studies. Data analysis was done using exploratory factor analysis (EFA) to identify the knowledge and skills obtained. Those are important to UiTM students' preparedness for the big data profession. From the exploratory factor analysis, 26 of the 40 items are included in the six constructs with factor loadings above 0.60: teamwork, student awareness and university readiness, programming language, student's effort, data storytelling and visualization, and data organization. These factors align with the finding made by [26], which identified the key competencies the employer needs for big data professions. In conclusion, higher education institutions need to focus on these skills in improving the existing program to meet better market demand and satisfy employer expectations since the score of factor loadings obtained are just satisfactory.**

Keywords*— *employability, big data analytics, BDA profession, competence, EFA

I. INTRODUCTION

Since 2009, Big Data has become one of the most widely discussed topics in many fields [1]. The global market for Big Data analytics is expanding rapidly, which is not surprising given the rising need. Therefore, data science has emerged as one of the top trending professions of the 21st century [2]. The term "data analytics" refers to mining large amounts of information for hidden patterns and insights through computational algorithms, programming, and statistical modelling.

As a means of capitalizing on big data's advantages, companies are putting in place the necessary infrastructure and analytics of big data and trying to train and educate their current workforce. However, companies are struggling to find enough qualified workers with the required big data capabilities to meet the rising demand. This has resulted in a severe shortage of skilled workers[3], and there is growing pressure on educational institutions to increase the number of students graduating prepared to fill this void [4]. Interestingly, a 2014 global survey of education institutions showed that a growing number of universities now offer a range of business intelligence, big data, and analytics-related courses for graduate programs, undergraduate programs and certificates [5].

Academically, data science has become the primary field responding to industrial demands for big data skills [6]. It is encouraging to see academic institutions trying to equip students with big data analytics skills. However, do the courses offered enough to prepare students with the knowledge and skills employers seek in new hires? This study aims to assess the Big Data analytics knowledge and skills obtained by students in higher education institutions. This study also attempts to investigate whether the courses learnt are sufficient to provide them with a comprehensive understanding of Big Data analytics.

II. LITERATURE REVIEW

A. Big Data Analytics

Big data analytics (BDA) uses cutting-edge analytical methods to sift through all that raw information and extract useful insights [7]. Real-time data that is too complicated and must be continually updated demands advanced analytical approaches which integrate data science, business intelligence and business analytics. Not only may big data analytics be used to discover previously unknown discoveries, but it can also rank or categorize the data according to how important the information it contains is [8]. BDA is necessary for businesses and society to get better results through intelligent decision-making, correct faulty data analysis, and lower risk and costs. To establish BDA, a company must be on interoperability, reusability, manageability, security, and maturity [9]. Companies that operate in high-pressure, ever-changing situations are the ones that have shown the most interest in big data analytics [10]. In today's business world, managers are increasingly making decisions based on the real-time insights provided by BDA, and they are also steering an increasing number of efforts in this direction [11].

B. Big Data Skills

Companies recognize data to be their asset and are investing heavily in the technologies and skills required for BDA [12]. An increasing gap exists between the need for and supply of qualified data analysts and scientists every year [13]. However, a lack of technical and analytical skills in the workforce remains a major obstacle as they explore or scale-out big data initiatives [14], particularly given that associations professions like data scientist and business analytics share common responsibilities [15].

Careers evolved [16] due to the demand for people with big data analytic skills [17]. Among the skills needed are Programming Languages (Python, JavaScript, SAS, SQL/NoSQL, MapReduce and Apache Spark), Machine Learning, Data Mining (RapidMiner, Apache Mahout and Knime), Predictive Analytics, Quantitative Analysis (SAS, IBM SPSS Statistics and the R language) and Data Visualization (Tableau and D3.js) [12].

References [18] and [19] examine the unique data analytics competencies that modern industry leaders feel accounting graduates should possess. According to [19], audit, accounting preparation, forensic accounting, and general accountancy consultancy are among the areas of accounting where big data and data analytics are starting to have an impact. Brink & Stoel [18] discover that communication and data interpretation skills are more valued than specific technical or statistical knowledge. Specifically, reference [19] identified four abilities for the efficient use of big data analytics. They included scepticism and questioning, critical thinking abilities, comprehension and capacity for analysis, and the ability to communicate findings. According to [20], employers are looking for people with solid functional and cognitive skills in data analytics, computing, and business, as well as a variety of social skills and particular personality qualities.

III. Research Methodology

A. Instrument Development

An adapted questionnaire was used as the research instrument. The questionnaire was adapted from three different sources, which are [21], [22], and [23]. This questionnaire consists of 40 items selected according to institution's suitability. The ruler and Options Scale [24] were used as a measuring scale in this questionnaire. The generated scale includes an option if the items are unrelated or inappropriate to the respondents. This study focused on the degree and master's students. The questionnaire has been disseminated online to many branches of computer science and mathematics students. It gathered 185 respondents.

B. Data Collection

This study focuses on bachelor's and postgraduate students at UiTM's Faculty of Computer and Mathematical Sciences. These students were chosen because they match the requirements needed to be considered for data scientist professions. They had formally taken big data courses during their studies.

Respondents comprised all degree students from the Faculty Computer and Mathematical Sciences across UiTM branches and students doing master's in data science at UiTM Shah Alam. The data collected were analyzed with the help of the Statistical Package for Social Sciences (SPSS). The exploratory factor analysis (EFA) was run to identify the number of factors relating to the quality of students in big data analytics skills.

C. Reliability and Validity Test

In this study, the content validity has been employed. The items was validated through the views of education experts in big data analytics. In addition, the construct was validated using the Exploratory Factor Analysis (EFA). As a result, the low loading and inability to meet the requirement of having a loading of more than 0.60 on a particular factor, a number of items in this study were removed. The reliability test used Cronbach's alpha to measure internal consistency. If the alpha value is higher than 0.9, the internal consistency is excellent, and if it is at least higher than 0.7, the internal consistency is acceptable [25].

In identifying the factors relating to the student's knowledge and skill sets in the big data profession, the items were selected according to the six constructs with factor loadings score was more than 0.60. From 40 items in the questionnaire, only 26 items were selected.

The reliability of all six factors was high (all at or above Cronbach's Alpha $\geq$ 0.830) indicating that the instrument items were highly internally accurate. The Cronbach's Alpha coefficient for each construct was shown in Table I. The result suggested that the internal consistency of the students' knowledge and skills sets in big data profession was high.

TABLE I. RELIABILITY TEST

Factor	Number of items	Cronbach's Alpha
1	6	0.931
2	5	0.901
3	3	0.888
4	4	0.929
5	5	0.906
6	3	0.830
Overall	26	0.946

IV. Results and Discussion

Participants in this study were 185 students whose majors included computer science and mathematics. Table II shows that 80.5% of the students in this study were female, while 19.5% were male. Moreover, most of the participating students came from the UiTM Kelantan Campus, which makes up 51.9%, followed by the UiTM Terengganu campus, with 36.76%.

TABLE II. RESPONDENTS' PROFILE

Variable	Category	N (185)	Percent (%)
Gender	Female	149	80.5
	Male	36	19.5
Part	1	1	0.5
	2	15	8.1
	3	27	14.6
	4	26	14.1
	5	58	31.4
	6	36	19.5
	7	21	11.4
	8	1	0.5
Campus	UiTM Kelantan	96	51.9
	UiTM Johor	2	1.10
	UiTM Negeri Sembilan	18	9.73
	UiTM Terengganu	68	36.76
	UiTM Shah Alam	1	0.54

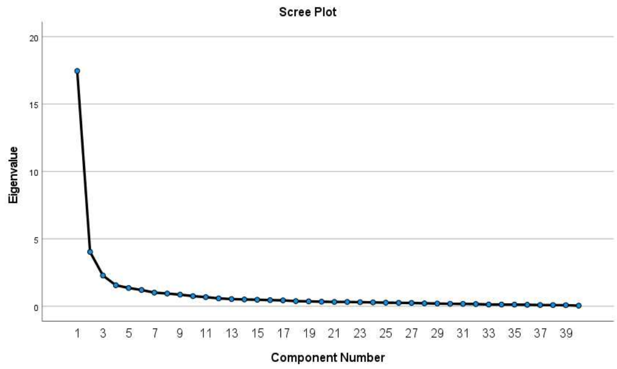

Fig 1. Scree Plot for Knowledge and Skill acquire by Students for Big Data Profession

As illustrated in Figure 1, there were six factors with eigenvalues greater than one. Scree plots are graphical representations used to find elbow shapes where only elements with eigenvalues greater than 1 are kept. The results of the exploratory factor analysis (EFA) for student's knowledge and skill sets in the big data profession is shown in Table III. The Cronbach Alpha value at 0.963 indicates the existence of excellent internal consistency. Alternately stated, a participant is more likely to respond positively to future survey questions if they respond positively to one [25].

The Principal Component Factor Analysis (PCA) with Varimax Rotation was performed for the 40 items of student's knowledge and skill sets in the big data profession. The factor of goodness of fit must be considered based on the Exploratory Factor Analysis (EFA) analysis to determine whether the analysis was sufficient, and the outcomes were reasonable. Bartlett's Test of Sphericity (0.000) value of less than 0.005 in Table III implies that the EFA run was satisfactory. There are enough samples to undertake factor analysis, as evidenced by the Kaiser-Mayer Olkin (KMO) test values over 0.6, which resulted in 0.935. The KMO value closer to 1.0 and the significance value of Barlett's Test close to 0.0 indicates the data at hand is adequate to proceed into Factor Analysis. The conducted EFA can therefore be considered as satisfactory because it satisfies the goodness of fit EFA.

TABLE III. EXPLORATORY FACTOR ANALYSIS FOR A STUDENTS' PREPAREDNESS IN BIG DATA ANALYTICS PROFESION

Items	Factors					
	1	2	3	4	5	6
Teamwork is essential in big data analytics careers.	0.665					
I enjoy working in a group.	0.795					
I am a good listener when working in a group.	0.723					
I am able to communicate my opinion.	0.635					
I know that a data scientist have to collaborate with workers from different departments.	0.726					
I know how to maintain good relationship with other team members.	0.738					
The university has computer laboratories for handling big data.		0.659				
The lecturers involved in teaching subjects on big data are knowledgeable in big data software.		0.768				
The lecturers involved in teaching subjects on big data have teaching skills in big data analytics.		0.739				
I am aware that there are job careers relating to big data.		0.658				
I realize that jobs in the field of big data analytics are highly demanded.		0.647				
I am good in Tableau Programming Language.			0.808			
I am good in Hadoop Programming Language.			0.863			
I am good in Spark Programming Language.			0.850			
I am interested to have a career in big data analytics.				0.718		
I am willing to upgrade myself in data communication skill.				0.752		
If I have the money, I am willing to spend on joining workshops to upgrade my data analytical skills.				0.831		

If I have the time, I am willing to upgrade my knowledge in big data analytic skills.				0.809		
I find it easy to interpret results from data analysis that I have conducted.					0.648	
I enjoy explaining results from data analysis to people.					0.734	
I know how to present research findings visually using graphs, tables and charts.					0.684	
I am not afraid of being questioned by the audience during the presentation.					0.718	
I enjoy describing data, how data were collected, and analysis done on the data.					0.715	
I am good in Data Visualization skills.						0.644
I am good in Data Wrangling and Preprocessing Skills.						0.605
I am good in Excel Programming Language.						0.607
Eigenvalue						
Percent Variance (%)	43.64	10.08	5.70	3.91	3.96	3.02
Cumulative Variance	43.64	53.72	59.42	63.33	66.73	69.75
Cronbach's Alpha	0.963 (number of items = 40)					
Barlett's Test of Sphericity (Chi-Square Value)	0.000 (6449.331)					
Kaiser-Meyer-Olkin (KMO) Measure of Sampling adequacy	0.935					

Six factors are involved in the students' for big data analytics knowledge and skills, as determined by the exploratory factor analysis in Table III. The first factor is teamwork, with loading factor values ranging between 0.635 to 0.795. The student understands the importance of teamwork in big data analytics. Thus, they enjoy working in groups, know how to maintain good relationships with other team members, express their ideas and listen to those of others, communicate their opinions, and actively listen to those of others. One transferable skill employers are looking for in graduates is the ability to work well in a team [26]. On the other hand, this finding is consistent with a study by [27], which found that teamwork and public presentation skills are the most crucial learning outcomes for future career opportunities. Project management and teamwork skills were frequently used in job advertisements and descriptions in big data professions namely, statistical analyst, big data analytics, data scientist, data analyst, and business analytics professionals [28]. From 2015 to 2018, the employer consistently advertised teamwork as a requirement to work for the company.

The student's awareness and university readiness in terms of resources and knowledgeable lecturers is the second factor that EFA addresses**.** To prepare the student for the job relating to big data analytics, it is essential to provide computer laboratories for handling big data. In addition, the lecturers who teach big data-related courses are well-versed in big data analytics and possess effective teaching skills. Therefore, university readiness aware students of the availability and high demand for careers in big data analytics. This factor's loading values range from 0.647 to 0.768. According to [29], universities should put efforts into helping students build their careers and train them to find work. In another study in Malaysia, this factor can be supported by [30]. They suggest that universities are supposed to and should be promoting employment by providing a learning environment that enables the development of skills, experiences, competencies, and values.

According to the EFA analysis, the programming language also influenced students' preparedness for big data analytics professions. The loading values for all items in this factor range between 0.808 and 0.863. Tableau, Hadoop and Spark Programming Language are the main components of the programming language factor. Furthermore, big data jobs often demand experience with various sophisticated analytical approaches, such as data mining, modelling, and advanced coding, as well as tools such as SAS, R, Hadoop, and SQL [31].

The EFA analysis also discovered that the students' efforts were a factor in students' preparedness for big data analytics professions. Students are prepared and willing to spend money on attending workshops to upgrade their analytical, as well as their knowledge of big data analytics and their data communications skills. Students are motivated to pursue professions in big data analytics, as well. This factor's loading values range from 0.752 to 0.831. This factor is supported by a study from [20] which found that employers seek workers with functional solid and cognitive competencies in data analytics with specific personality traits. Additionally, employers understand that graduate students hardly ever have the necessary skills and are ready to provide more training. The employer's intentions to provide more training are aligned with the student's willingness to do so.

Another factor in students' preparedness for employment in big data analytics through EFA was identified as data storytelling and data visualization. All loading factors for this factor fall within the range of 0.648 to 0.734. The students find it easy to interpret results from data analysis conducted, enjoy explaining results from data analysis to people, and know how to present research findings visually using graphs, tables, and charts. In addition, students aren't afraid of being questioned by the audience during the presentations and, lastly, enjoy describing data, how data were collected, and analysis chosen on the data. A similar study about big data professions by [28] explains that data analysts are expected to apply their communication skills and deliver speaking, written, and visual presentations of pertinent information to the target audiences.

Data organization, the sixth factor, and good data visualization skills and excel programming language. To make unusable raw data, data wrangling is primarily used and important in students' preparedness for big data analytics professions. For this factor, all loading factors fall between 0.605 and 0.644. By highlighting the significance of data exploration, handling missing values, reshaping data, and filtering data, [32] demonstrates the steps involved in data wrangling in phyton. Once the raw data has been transformed into an effective dataset, it can then be used for data visualization, data analysis, and model training.

Table IV maps the knowledge and skills related to the big data analytics acquired by students and the employer's view. Employers are looking for someone with the following four competencies: data analytics, computing, business and soft skills [26]. Out of four competencies, business is excluded since it does not match with the items.

TABLE IV. MAPPING OF KNOWLEDGE AND SKILLS WITH THE BIG DATA COMPETENCIES

Knowledge and skill	Data Analytics	Computing	Soft skill
Teamwork			/
Student awareness and university readiness	/		/
Programming Language		/	
Student's effort			/
Data Storytelling and visualization	/		
Data Organization	/		

V. CONCLUSION

In this paper, we identify the knowledge and skill acquired by the students who had formally taken big data courses during their studies. The questionnaire consists of 40 items and has been distributed to 185 students. Students are required to evaluate themselves to answer the questions. Exploratory factor analysis has been used to analyze the items. Of the 40 items, 26 are included in the six constructs with factor loadings scores above 0.60. The constructs are teamwork, student awareness and university readiness, programming language, student effort, data storytelling and visualization, and data organization. We can conclude that the constructs are the knowledge and skills obtained throughout their study.

This finding aligns with the finding by the author [26] and tries to seek whether big data programs enable students to acquire the competencies required by employers. This is because, from the finding, the students are ready to fulfil the requirement needed by the employer. The students also stated the university is preparing them for the big data profession. It can be concluded our course and program offered are relevant and meet the industry demand. In addition, [20], who identified the skills set for big data analytics, also recommends that higher education institutions need to adjust their curriculum to satisfy industry demand better. Future studies should include business as one of the key competencies evaluated. In addition, more respondents are required to reveal adequate information.

ACKNOWLEDGMENT

The registration fee is funded by Pembiayaan Yuran Prosiding Berindeks (PYPB), Tabung Dana Kecemerlangan Pendidikan (DKP), Universiti Teknologi MARA (UiTM), Malaysia.

REFERENCES

[1] M. A. Waller and S. E. Fawcett, "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management," *J. Bus. Logist.*, vol. 34, no. 2, pp. 77–84, 2013, doi: 10.1111/JBL.12010.

[2] T. H. Davenport and D. J. Patil, "Data scientist: The sexiest job of the 21st century," *Harv. Bus. Rev.*, 2012.

[3] R. H. . Chiang, P. Goes, and E. . Stohr, "Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline," *ACM Trans. Manag. Inf. Syst.*, vol. 3, no. 3, pp. 1–13, Oct. 2012, doi: 10.1145/2361256.2361257.

[4] T. Davenport, *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, 2014.

[5] B. Wixom, T. Ariyachandra, D. Douglas, M. Goul, and B. Gupta, "The current state of business intelligence in academia: The arrival of big data," *Commun. Assoc. Inf. Syst.*, vol. 34, no. 1, p. 1, 2014.

[6] I. Song and Y. Zhu, "Big data and data science: what should we teach?," *Expert Syst.*, vol. 33, no. 4, pp. 364–373, 2016.

[7] S. Tiwari, H.-M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Comput. Ind. Eng.*, vol. 115, pp. 319–330, 2018.

[8] A. Oussous, F. Z. Benjelloun, and A. A. Lahcen, "Big Data technologies: A survey," *J. King Saud Univ. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, 2018.

[9] N. Das, L. Das, S. . Rautaray, and M. Pandey, "Big data analytics for medical applications," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 2, pp. 35–42, 2018, doi: 10.5815/ijmecs.2018.02.04.

[10] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: Certain investigations for research and applications," *Int. J. Prod. Econ.*, vol. 176, pp. 98–110, Jun. 2016, doi: 10.1016/J.IJPE.2016.03.014.

[11] I. D. Constantiou and J. Kallinikos, "New Games, New Rules: Big Data and the Changing Context of Strategy:," *J. Inf. Technol.*, vol. 30, no. 1, pp. 44–57, Mar. 2015, doi: 10.1057/JIT.2014.17.

[12] A. Abbasi, S. Sarker, and R. H. L. Chiang, "Big data research in information systems: Toward an inclusive research agenda," *J. Assoc. Inf. Syst.*, vol. 17, no. 2, p. 3, 2016.

[13] CrowdFlower, "DATA SCIENTIST 2017 BROUGHT TO YOU BY," 2017.

[14] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, "Analytics: The real-world use of big data," *IBM Corporation,Somers, New York*, 2012.

[15] R. Agarwal and V. Dhar, "Big data, data science, and analytics: The opportunity and challenge for IS research," *Information systems research*, vol. 25, no. 3. INFORMS, pp. 443–448, 2014.

[16] S. Levine, "'8 things you can do to evolve your career.'"

[17] J. Hardin *et al.*, “Data science in statistics curricula: Preparing students to ‘think with data,’” *Am. Stat.*, vol. 69, no. 4, pp. 343–353, 2015.
[18] W. D. Brink and M. D. Stoel, “Analytics knowledge, skills, and abilities for accounting graduates,” in *Advances in accounting education: Teaching and curriculum innovations*, Emerald Publishing Limited, 2019.
[19] K. Mcbride and C. Philippou, “‘Big results require big ambitions’: big data, data analytics and accounting in masters courses,” *Account. Res. J.*, 2021.
[20] A. Persaud, “Key competencies for big data analytics professions: a multimethod study,” *Inf. Technol. People*, 2020.
[21] K. E. Arnold, S. Lonn, and M. D. Pistilli, “An exercise in institutional reflection: The learning analytics readiness instrument (LARI),” in *Proceedings of the fourth international conference on learning analytics and knowledge*, 2014, pp. 163–167.
[22] M. Oster, S. Lonn, M. D. Pistilli, and M. G. Brown, “The learning analytics readiness instrument,” in *Proceedings of the sixth international conference on learning analytics & knowledge*, 2016, pp. 173–182.
[23] “Unlocking Big Data: Business Value, Drivers and Challenges,” *IBM Corporation,Somers, New York*, 2022. http://unlockingbigdata.com/
[24] R. Yusoff and R. Mohd Janor, “Generation of an interval metric scale to measure attitude,” *SAGE Open*, vol. 4, no. 1, 2014, doi: 10.1177/2158244013516768.
[25] N. Blunch, *Introduction to structural equation modeling using IBM SPSS statistics and AMOS*. Sage, 2012.
[26] M. Humburg, R. Van der Velden, and A. Verhagen, “The employability of higher education graduates,” *Maastricht Publ. Off. Eur. Union*, p. 4, 2013.
[27] M. J. Haigh and M. P. Kilmartin, “Student perceptions of the development of personal transferable skills,” *J. Geogr. High. Educ.*, vol. 23, no. 2, pp. 195–206, 1999.
[28] M. A. Halwani, S. Y. Amirkiaee, N. Evangelopoulos, and V. Prybutok, “Job qualifications study for data science and big data professions,” *Inf. Technol. People*, 2021.
[29] D. Cole and M. Tibby, “Defining and developing your approach to employability: A framework for higher education institutions,” *Heslingt. High. Educ. Acad.*, 2013.
[30] I. A. G. Azmi, R. C. Hashim, and Y. M. Yusoff, “The employability skills of Malaysian university students,” *Int. J. Mod. Trends Soc. Sci.*, vol. 1, no. 3, pp. 1–14, 2018.
[31] K. Ahern and N. Keller, “‘Are you a big data professional?,’” *Marketing News*.
[32] B. Priyadharshini, “Let’s Understand All About Data Wrangling!,” *Analytics Vidhya*, 2021.