Students' Views and Difficulties in Data Analysis at the Institute of Adult Education in Tanzania

Kija Steven Magembe
Institute of Adult Education, Morogoro-Campus
Contact address: 827 Morogoro, Tanzania
E-mail: *stevenkmagembe2001@gmail.com* or smagembe@iae.ac.tz

Abstract

The purpose of this study was to examine undergraduate research projects completed in 2019/2020 academic year at the Institute of Adult Education in terms of the types of statistical techniques commonly used and difficulties encountered by students in data analysis process. The subjects in this study were the finalist undergraduate students of 2019/2020 comprised 125 respondents. The study adopted a mixed methods approach. Sampling techniques adopted in this study was purposive. Data collection techniques in this study were questionnaires, focus group discussions and interviews. Both quantitative and qualitative data analyses were involved. Descriptive statistics and t-test analysis technique were used for data analysis. Students' difficulties of data analysis indicated that, sixty-nine (69%) had a problem with statistical hypothesis testing; sixty five percent (65%) had a problem with analysis of qualitative data. Eighty one percent (81%) of the participants had problem with choosing the right statistical test for data analysis. In addition, descriptive statistics were mostly used (89.6%) and it was indicated that there was a significant difference between the problems experienced by male and female students ($p<0.05$). The themes that emerged from the data analysis were analytical skills, numerical skills and technical skills. It would be beneficial for students at the Institute of Adult Education to receive trainings on data analysis skills. Findings from this study emphasize the imperative to understand the challenges these students face as they continue their educational journeys in Higher Learning Institutions in Tanzania.

Keywords: *Data analysis, Views, Difficulties, Undergraduate Students*

Introduction

Research activity is considered as one of the high-impact educational practices in that the vital skills and attitude for lifelong learners can be cultivated through inquiry (Hunter, Laursen, & Seymour, 2007). Walliman (2011) suggests that research is a very general term for an activity that involves finding out, in a more or less systematic way things you did not know. A more academic interpretation is that research involves finding out about things that no one else knew either. It is about advancing the frontiers of knowledge. Undergraduate research is defined as any teaching and learning activity in which undergraduate students are actively engaged with the research content, process or problems of their discipline (Zimbardi & Myatt, 2014). In addition, undergraduate research-based projects are capstone experiences that provide students with an opportunity to answer a research question within a disciplinary framework under supervision (Ashwin, Abbas & McLean, 2016). They form an essential component of many undergraduate degrees, provide a transition between course work and independent research and may result in publishable research. That is, research is not merely a pursuit of academic career and advancement of knowledge (i.e., content) but also an aspect of the learning process (Willison & O'Regan, 2007). Moreover, developing research skills is essential in adult education. According to Bandele (2004), educational research is carried out, to increase human knowledge to solve contemporary problems and establish decision-making basis, to make discoveries and promote the contemporary modern evidence to help educational innovations and improve educational services. In research, once data is collected, the next step is to get insights from it. Data analysis is perhaps the most important component of research.

According to Marshall & Rossman (2016), data analysis can be defined as the process of bringing order, structure, and interpretation to the mass of collected data. While Hatch (2002 states that data analysis is a systematic search for meaning. Data analysis is defined as a process of cleaning, transforming, and modelling data to discover useful information for business decision-making. The purpose of data analysis is to extract useful information from data and take decisions based upon the data analysis. Data analysis is the process of systematically applying statistical or logical techniques to describe and illustrate, condense, recap and evaluate data. Analysis is the procedure to make broad generalizations by identifying trends and situations (phenomena) among present information (UNESCO, 2018).

Analysis means organizing and interrogating data in ways that allow researchers to see patterns, identify themes, discover relationships, develop explanations, make interpretations, mount critiques, or generate theories (IEE, 2010). It often involves synthesis, evaluation, interpretation, categorization, hypothesizing, comparison, and pattern finding. The purpose of the data analysis in research studies is to find answers to the research question and to facilitate the researcher to determine the trends and inter-relationship of different variables included in the study (Sham & Resnik, 2003). Writing an academic report is one of the requirements every undergraduate student has to accomplish in completing his/her degree program; it is written based on research. An objective of the research-based project that is highly regarded at the Institute of Adult Education is to develop students' research skills. Students can choose a project on any topic on various aspects necessarily within their area of study. Each project is undertaken individually over a two-semester period supervised by an academic staff member. Hence, statistics has become a substantial part of the project. However, students at the Institute of Adult Education have no module in statistics. The main concern is whether what they have learned in the modules is enough to equip them well for using statistical techniques in data analysis, and what difficulties they encountered. The study is based on a survey that was conducted amongst the final year undergraduate students in the academic year 2019/2020 upon completion of their projects. A questionnaire was designed to find out students' views and difficulties in data analysis and also the questionnaire was used to identify the common data analysis techniques which were used by the students at the Institute of Adult Education in Tanzania. One major area receiving less attention when examining problems facing Universities and Institutions in Africa, including Tanzania, are those difficulties encountered by undergraduate and postgraduate students in data analysis and this is compounded by several factors (Duze, 2010; World Economic Forum, 2021). Some of these challenges includes; problems with choosing the right statistical test for data analysis, lack of writing skills, lack of training in research work, poor knowledge in using the statistical package, lack of mathematics and statistics course and the problem with the analysis of qualitative data. Many of these problems are shared more by undergraduate students, but what are the common problems/difficulties facing these students? What are data analyses techniques are commonly used by students in their research-based projects

at the Institute of Adult Education? Are the problems faced by male students significantly different from those encountered by female students? The institute must take significant steps to identify these problems so as to either eliminate them or at least minimize them. The purpose of this study was to find out students' views of data analysis, statistical techniques used and difficulties encountered during data analysis process and compare the difficulties experienced by the female students to their male counterparts at the Institute of Adult Education, Tanzania.

To address these aims, and to answer the question posed by the title of this study, the following research questions were proposed:

i) What difficulties do undergraduate students experience in data analysis process?

ii) What are the data analyses techniques which are commonly used by students in their research-based project at the Institute of Adult Education?

iii) Are the problems faced in data analysis by male students is significantly different from those encountered by female students?

Literature Review

In recent decades, statistics educators have made much progress in determining what students need to know and how best to develop their abilities in the domain of data analysis. Data analysis tasks prove most successful when students engage deeply with the data and are involved with the planning and development of the analysis. Despite or perhaps even because of the wide range of applications of data analysis and statistics, learning these concepts in context can be difficult for students. As Moore (1990) explains, "data are not merely numbers, but numbers with a context" (p. 96). Interdisciplinary applications of concepts in statistics provide students with many opportunities to engage with these ideas, yet the contextual nature of the data means that each application will be unique thus presenting its challenges for students. Educators should strive to help students to see the connections between context-specific applications of data analysis and the big ideas and concepts that make up the discipline of statistics, but before we can do that we need to identify exactly how students engage with data analysis concepts in applied settings. This case study intends to contribute to that understanding.

Process Challenges

Process challenges are the group of challenges encountered while processing and analyzing the data. Process challenges involves capturing the data, interpreting and presenting the results. As large datasets are usually non-relational or unstructured, thus processing such semi-structured data sets at scale poses a significant challenge; possibly more than managing big data (BD) (Kaisler, Armour, Espinosa & Money, 2013). Several data processing related challenges can be grouped into 4 steps: data mining and cleansing, data aggregation and integration, data analysis and modelling and data interpretation.

Data Mining and Cleansing:

This challenge relates to extracting and cleaning data from a collected pool of large unstructured data. Advocates of BD and BDA perceive that identifying a better way to mine and clean the BD can result in big impact and value (Chen, Chen, Jiang, Ooi, Shi, Vo & Wu, 2012). Due to its strident, vibrant, diverse, inter-related and unreliable features, the mining, cleansing and analysis confirm to be very challenging (Chen, Chen, Du, Li, Lu, Zhao & Zhou, 2013). For instance, in the UK National Health Service (NHS) there are many millions of patients' records comprising of medical reports, prescriptions, x-ray data, etc. Physicians make use of such data – if for instance incorrect information is stored this may lead to physicians wrongly diagnosing conditions, resulting in inaccurate medical records. However, to make use of this huge data in a meaningful way, there is a need to develop an extraction method that mines out the required information from unstructured BD and articulate it in a standard and structured form that is easy to understand.

Data Aggregation and Integration:

This process challenge relates to aggregating and integrating clean data mined from large unstructured data. BD often aggregates varied online activities such as tweets–retweets, microblogging, and likes on Facebook that essentially bear diverse meanings and senses (Edwards & Fenwick, 2015). This characteristically amorphous data naturally lacks any binding information.

Data Analysis and Modelling

Once the data has been captured, stored, mined, cleaned and integrated, it comes for the data analysis and modelling for BD. Outdated data analysis and modelling centres around solving the intricacy of relationships between schema-enabled data. As BD is often noisy, unreliable, heterogeneous, and dynamic; in this context, these considerations do not apply to non-relational, schema-less databases (Shah, Rabhi & Ray, 2015). From the perspective of differing between BD and traditional data warehousing systems. Kune, Konugurthi, Agarwal, Chillarige & Buyya (2016) report that although these two have similar goals; to deliver business value through the analysis of data, they differ in the analytics methods and the organization of the data. Consequently, old ways of data modelling no longer apply due to the need for unprecedented storage resources/capacity and computing power and efficiency (Barbierato, Gribaudo & Laoono, 2014). Thus, there is a need for new methods to manage BD for maximum impact and business value. It is not merely knowing about what is currently trendy, but also need to anticipate what may happen in the future by appropriate data analysis and modelling (Chen et al., 2013).

Data Interpretation

This step is relatively similar to visualizing data and making data understandable for users that is the data analysis and modelling results are presented to the decision-makers to interpret the findings for extracting sense and knowledge (Simonet, Fedak & Ripeanu, 2015). The astounding growth and multiplicity of unstructured data have intensely affected the way people do process and interpret new knowledge from these raw data. As much of these data both instigate and reside as an online resource, one open challenge is defining how Internet computing technological solutions have evolved to allow access, aggregate, analyse, and interpret BD. Another challenge is the shortage of people with analytical skills to interpret data (Phillips-Wren & Hoskisson, 2015).

Materials and Methods

In this study, the researcher designed a mixed-method. Mixed methods involved the use of both quantitative and qualitative methods in a single study. The research was conducted at the Institute of Adult Education. The population under this study comprised 125 participants. Sampling techniques adopted in this study was purposive. Data collection techniques in this study were questionnaires,

focus group discussions (FGDs) and interviews. The questionnaire that was used for knowing a research question the researcher decided to use closed-ended questionnaires using a Likert Scale. The Likert Scale is the most common scale that is used in measuring attitude, opinion and perception of respondents towards the subject. The instruments were pre-tested before collecting the data. To ensure that the instrument for data collection was valid and reliable, the instrument was tested for face and content validity.

A focus group interview form developed by the researcher was used to collect qualitative data. A focus group is a qualitative technique that emphasizes dynamic group interaction and provides specific information on a selected topic in a relatively short period (Vaughn, Schumm & Sinagub, 1996). Focus group interviews are based on the principle of interaction among group members. In the development of this focus group interview form, literature was reviewed in detail and four questions were prepared to determine problems faced by undergraduate students in data analysis. Then, three experts' opinions were taken, the draft form was examined by two students and the questions were revised accordingly. The final form of the focus group interview was conducted with eight students and it took approximately 90 minutes. The English language was used for administering the survey scale and focus group interview.

The data collected were analysed using descriptive statistics–frequency, percentages, means and standard deviation to answer the research questions. Quantitative data were entered and analysed using SPSS (Statistical Package for Social Science) version 24. The t-test was used to test the hypotheses formulated at the 0.05 level of significance. A thematic analysis technique was used in the analysis of qualitative data. Each transcript was analysed individually and thematically. The participants' views were made sense of by developing an interpretive relationship with the transcripts. As the analysis developed, the researcher began to look for patterns in the codes, called themes. By dividing the text into meaning units/themes, the researcher was able to identify commonalities, differences, and contradictions across all the participants describing the same or different phenomena and finally transcribed and coded the interviews in considerable detail in an attempt to make sense of every individual's experience.

Results

Data Analysis Difficulties

In order to determine the views of students about difficulties encountered in data analysis process, a survey was conducted; the mean and standard deviation values of the items in the survey are presented in Table 1.

Table 1: STUDENTS' VIEWS OF DIFFICULTIES ENCOUNTERED IN DATA ANALYSIS PROCESS (N=125)

	SD f (%)	D f (%)	N f (%)	A f (%)	SA f (%)	Mean	SD
1. I have difficulty in familiar with types and level of variables	10(8)	15(12)	8(6.4)	40(32)	52(41.6)	2.89	1.24
2. I have a problem with analysis of qualitative data	5(4)	3(2.4)	12(9.6)	40(32)	65(52)	2.92	1.22
3. Poor skills in data analysis	10(8)	25(20)	6(4.8)	40(32)	44(35.2)	2.87	1.24
4. Data cleaning/preparing data for analysis is difficult	12(9.6)	12(9.6)	13(10.4)	36(28.8)	52(41.6)	2.57	1.12
5. I have a problem with choosing the right statistical test for data analysis	12(9.6)	6(4.8)	5(4)	40(32)	62(49.6)	3.46	0.13
6. I have a problem with statistical hypothesis testing	6(4.8)	11(8.8)	9(7.2)	30(24)	69(55.2)	3.78	1.02
7. Lack of training in research work	8(6.4)	13(10.4)	5(4)	43(34.4)	56(44.8)	3.54	1.11
8. Poor knowledge in using statistical package eg. SPSS and EXCEL	7(5.6)	23(18.4)	8(6.4)	42(33.6)	45(36)	3.61	0.21
9.Collecting Poor quality of data	10(8)	10(8)	5(4)	40(32)	60(48)	5.02	1.15
10. I have less experience in research works	7(5.6)	4(3.2)	4(3.2)	50(40)	60(48)	3.76	1.12

Source: Survey Data, 2019.

Note: SA=Strongly Agree, A=Agree, N= Neutral, D=Disagree and SD=Strongly Disagree

As seen in Table 1, the means of items about students' views vary between 2.57 and 5.02. Table 1 also reveals that sixty-two per cent (62%) of the students, had a problem with data analysis and interpretation of the findings, sixty-nine (69%) had a problem with statistical hypothesis testing, sixty-five per cent (65%) had a problem with the analysis of qualitative data. Eighty-one per cent (81%) of the participants had a problem with choosing the right statistical test for data analysis. Furthermore, some respondents claimed to collect data of poor quality (48%).

Analysis of Focus Group Discussions

To supplement the results and to fill the gaps left in the questionnaire, the qualitative approach was used. This kind of technique looks more likely to give more substance to reveal detailed information. Qualitative research offers unique opportunities for understanding complex, nuanced situations where interpersonal ambiguity and multiple interpretations exist. This methodology is used to get information about how people think, feel and act and what they know. This section of the research was conducted using focus group discussions (FGDs) consisting of 10 students. The main goal of a focus group is to find answers to the "why", "what", and "how" questions. The information collected was presented in a narrative form that includes the description and analysis of data.

In the focus group interview, students stated that they have difficulty in data analysis due to having poor statistics skills. Relating to basic statistical skills, they stated that they have difficulty in understanding what they read and have to read many times to understand. In addition, some participants pointed out that they also have a problem in mathematics. Three themes along with four sub-themes were identified from some of the participants (*Table 2*).

Table 2: PARTICIPANTS VIEWS ON THE DIFFICULTIES OF DATA ANALYSIS

Themes	Sub-themes	Representative Quotes
Analytical skills	Problems with basic statistical skills	1. I am having problems with basic statistical skills. In addition, to interpret and apply statistics in the research study is a problem (S7). 2. My problem is difficulty in understanding statistics concepts; sometimes it becomes difficult for me to do analysis. (S8).
Numerical skills	Poor mathematics backgrounds	I am having the most trouble in data analysis because I did not get a Mathematics course at the Institute. (S5).
Technical skills	1. Problems in analysis of qualitative data 2. Problem in data visualization	1. I have problems in the analysis of qualitative data. For example, when I have open-ended questions, it is difficult for me to analyze the data. (S2). 2. When you have collected, the final challenge now is how to make sense of the data you have collected. (S3). 3. We are trying to analyze data ourselves, but we get lost, sometimes we hire a specialist from outside to analyze our data, although it is strictly prohibited (S10).

Data Analysis Techniques Used by Students in their Research-based Projects

As illustrated in Table 3, 112 (89.6%) of 125 research-based project reports, descriptive statistics were used, in 5 reports, inferential statistics (4 %), the parametric statistical technique was used at least once (0.8%), a non-parametric statistical technique was not used at all and in 7 reports (5.6%), a content/ thematic analysis was used.

Table 3: DISTRIBUTION BY DATA ANALYSIS TECHNIQUES

Data analysis techniques	frequency	Percentage (%)
Descriptive statistics	112	89.6
Inferential statistics	5	4.0
Parametric	1	0.8
Non-parametric	0	0.0
Thematic/Content analysis	7	5.6
Total	**125**	**100**

Source: Survey Data, 2019

Problems Encountered by Male and Female Students (N = 125)

Table 4 shows the computed t-test analysis of the data from the responses of all-male respondents (male undergraduate students) and all-female respondents (female undergraduate students).

Table 4: SUMMARY TABLE OF T-TEST ANALYSIS OF THE PROBLEMS ENCOUNTERED BY MALE AND FEMALE STUDENTS (N = 125).

Source	N	Mean	SD	t-calc.	Df	Alpha	Critic- t	Value
Male students	45	3.19	0.45	6.88	123	0.05	1.96	Sig.
Female students	80	3.57	0.63					

Source: Survey Data, 2019

Table 4 shows the mean scores of the problems faced by female students to be 3.57 with a SD of 0.63 and that for the male students to be 3.19 with a SD of 0.45. This implied that the mean score of the problems experienced by female students was slightly higher than that of the males. When the data for the two groups were subjected to t-test analysis at the 0.05 level of significance, the

results showed a computed t of 6.88 which was higher than the table value of 1.96. Since the obtained value was higher than the critical table value, it was concluded that there was a significant difference between the problems experienced by male and female students at the institute. The mean difficulty in female was significantly higher in female than in male, $t(123) = 6.88$, $p<.001$.

Discussions

The purpose of this study was to find out students' views of data analysis, statistical techniques used and difficulties encountered in the data analysis process and compare the difficulties experienced by the female students to their male counterparts at the Institute of Adult Education, Tanzania. The quantitative results indicated that data analysis is not an easy task for undergraduate students; many students do not know how to choose the right statistical test. The study findings enhance our understanding that data analysis involves a synthesis of a complex range of analytical and rhetorical skills as well as academic writing skills and an understanding of what is meant by critical analysis and argument. This idea is supported by other studies e.g., Safari, Navazeshkhah, Azizi, Ziaei, & Sharafi (*2015*) who stated that lack of research experience affects the level of interest and participation of students in research activities. Research experience improves the interest and skill of the students in conducting research activities and it can be achieved by participation in workshops related to research. This finding is parallel with the opinion of the students in this study (48%). But few students (8.8%) confessed that they had experience in research works.

Research is a complex skill that requires different techniques to use the database and assess the results of research. Others need planned thinking, such as knowing which tools are effective and appropriate for different research questions. In addition, to develop the proficiency and development capabilities in all these areas, students must be capable of integrating all of these (Wietse-vander, Anouke, Anje, Douwe & Marc, 2012). Rani & Priya (2014) as cited by Zain (2020) pointed out that students face difficulties due to a lack of research methods and time. To select the appropriate statistical method, one needs to know the assumption and conditions of the statistical methods so that proper statistical methods can be selected for data analysis (Nayak & Hazra, 2011). Other than knowledge of the statistical methods, another very important aspect is the nature and type of the data collected and the objective of the study because as per objective, corresponding statistical methods are selected which are suitable on given data.

While selecting appropriate statistical techniques, types and levels of variables need to be determined correctly. The present study indicated that a descriptive statistics technique was mostly used by students (89.6%). Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

Elmore & Woehlke (1988) reported that descriptive statistics is frequently used in journal articles in education in the United States. Likewise, Baumberger & Bangert (1996) reviewed articles in the Journal of Learning Disabilities in terms of research designs and statistical techniques, they found that descriptive statistics, t-test, one-way ANOVA, Pearson Product-Moment Correlation Coefficient, Chi-square, which they referred to as simple statistical techniques, were most frequently used. Bangert & Bamberger (2005) stated that descriptive statistics and Pearson Product-Moment Correlation Coefficient were frequently used in the articles published in the Journal of Counselling and Development over 11 years. Good descriptive research relies primarily on low-inference, low-assumption methods that use no or minimal statistical adjustments. Measures of central tendency, variation, and basic frequency analyses are particularly useful tools. Although there can be no claim of causation, any descriptive finding that uncovers a socially relevant "truth" in the data can serve as the foundation for generating hypotheses, prioritizing possible causal mechanisms, or otherwise pointing toward causal understanding (Romero & Ventura, 2012).

The selection of the appropriate statistical methods is very important for quality research. It is important that a researcher knows the basic concepts of the statistical methods used to conduct research study that would produce a valid and reliable results. There are various statistical methods that can be used in different situations. Each test makes particular assumptions about the data. These assumptions should be taken into consideration when deciding which the most appropriate test is. Wrong or inappropriate use of statistical methods may lead to defective conclusions, finally would harm the evidence-based practices. An essential component of ensuring data integrity is the accurate and appropriate analysis of research findings. Improper statistical analyses distort scientific findings, mislead casual readers (Shepard, 2002), and may negatively influence the public perception of research. Hence, an adequate knowledge of statistics and the appropriate use of statistical tests are important for improving and producing quality research.

The challenges faced by finalist-undergraduate students at the Institute of Adult Education were identified. The study reveals that students have difficulty in data analysis and a descriptive statistics technique was mostly used by students when analysing their data. It was concluded that there was a significant difference between the problems experienced in data analysis by male and female students. Selection of the appropriate statistical methods is very important for the quality research. It is important that a researcher knows the basic concepts of the statistical methods used to conduct research study that can produce a valid and reliable results. There are various statistical methods that can be used in different situations. Each test makes particular assumptions about the data. These assumptions should be taken into consideration when deciding which the most appropriate test is. Wrong or inappropriate use of statistical methods may lead to defective conclusions, finally would harm the evidence-based practices. Hence, an adequate knowledge of statistics and the appropriate use of statistical tests are important for improving and producing quality research.

The themes that emerged during the Focus Group Discussions were classified under three perspectives-analytical skills, numerical skills, and technical skills which were further categorized into subthemes. The analytical skills challenges were categorized into subtheme- problems with basic statistical skills. The technical skills challenges were divided into two subthemes- problems in the analysis of qualitative data, and problems in data visualization challenges. Almost all the researchers have a consensus that it is the need of the day to address these challenges at the respective levels to make the higher learning institutions in the true sense. The institutions administration needs to take initiatives for resolving the problems faced by students during the learning process.

Experience has shown that students tend to encounter difficulties in data analysis when carrying research projects. Although the project is not necessarily a statistical project, many of these projects involve data analysis. However, students have a limited understanding in data analysis, thus the main concern is whether they equip themselves in data analysis and what difficulties they encounter while handling data. Based on these findings, the following recommendations could be made. Fields within educational sciences should make extra effort to equip students with sufficient knowledge of statistical techniques. Courses of "Statistics and Research Methods" should be introduced at the institute and it should be taught in undergraduate programs and students should be provided with basic knowledge of statistical techniques. Undergraduate research methods courses

can be considered as an effective way to increase these research skills (Pettijohn, Naylor & Piroch, 2013). As data analysis has become such a critically valuable skill in so many areas of the world, statisticians will have to think harder about what makes for a good data analyst. Fourthly, we need to develop better ways to train analysts to do the right thing. Learning by doing will always be a critical aspect of data analytic training, if only because practice is essential. It would be beneficial for supervisors working at the institute to receive training on how to equip future researchers with the necessary data analysis skills and students must have the necessary skills to analyse the data, getting trained to demonstrate a high standard of research practice. Ideally, researchers must possess more than a basic understanding of the rationale of selecting one statistical method over the other to obtain better data insights.

References

Ashwin, P., Abbas, A., & McLean, M. (2016). How does completing a dissertation transform undergraduate students' understandings of disciplinary knowledge? Assessment Evaluation Higher Education, 42(4), 517–530. https://doi.org/10.1080/ 02602938.2016.1154501

Bandele S.O. (2004). Educational Research in Perspective, Ibadan: Niyi. Retrieved from https://files.eric.ed.gov/fulltext/ED567134.pdf

Bangert, A. W., & Baumberger, J. P. (2005). Research and statistical techniques used in the Journal of Counselling and Development: 1990–2001. Journal of Counselling and Development, 83(4), 480–487. https://doi.org/10.1002/j.1556-6678.2005.tb00369.x

Barbierato, E., Gribaudo, M. S., & Laoono, M. (2014). Performance evaluation of NoSQL big-data applications using multi-formalism models. Future Generation Computer Systems, 37(3), 345–353. https://doi.org/10.1016/j.future.2013.12.036

Baumberger, J. P., & Bangert, A. W. (1996). Research design and statistical techniques used in the Journal of Learning Disabilities, 1989-1993. Journal of Learning Disabilities, 29(3), 313-316.

Chen, G., Chen, K., Jiang, D., Ooi, B.C., Shi, L., Vo, H. T., & Wu, S. (2012). E3: An elastic execution engine for scalable data processing. Journal of Information Processing, 20 (1), 65-76. https://doi.org/10.2197/ipsjjip.vol.20.65

Chen. J., Chen, Y. Du. X. Li, C. Lu. J. Zhao. S., & Zhou, X. (2013). Big data challenge: a data management perspective. Frontiers of Computer Science, 7(2), 157–164. https://scholar.google.com/citations?user=25xBm5EAAAAJ&hl=en

Duze, C. O. (2010). An Analysis of Problems Encountered by Postgraduate Students in Nigerian Universities, Journal of Social Science, 22(2), 129-137. https://doi.org/ 10.1080/09718923.2010.11892793

Edwards, R., & Fenwick, T. (2015). Digital analytics in professional work and learning. Studies in Continuing Education, 1(1), 1–15. https://doi.org/10.1177/14749041156

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher and Review of Educational Research from 1978 to 1987. Educational Researcher,

17(9), 9–20.

Hatch, J. A. (2002). Designing Qualitative Research in Education Settings. New York: State University of New York Press. 308pp.

Hunter, A., Laursen, S. L., & Seymou, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. Science Education, 91(1), 36–74. https://doi.org/10.1002/sce.20173

IEE (International Encyclopaedia of Education) (2010). Retrieved from https://www.sciencedirect.com/ topics/social-sciences/data-analysis.

Kaisler S., Armour F., Espinosa J. A., & Money, W. (2013), data: Issues and challenges moving forward 46th Hawaii International Conference on System Sciences (HICSS) pp.995-1004. https://doi.org/10.1109/HICSS.2013.645

Kune R., Konugurthi P. K., Agarwal A., Chillarige, R. R., & Buyya, R.(2016). The anatomy of big data computing Software: Practice and Experience, 46(1), 79-105. https://doi.org/10.1002/spe.2374

Loi, S. L., & MacLaurin, T. L. (2000). Using technology to support teaching statistics to a large class of business students (abstract). In Proceedings TIME 2000, the International Conference on Technology in Mathematics Education. pp.288.

Marshall, C., & Rossman, G. (2016). Designing Qualitative Research. 6th Edition, SAGE, Thousand Oaks. Thousand Oaks, California: SAGE. 323pp.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), On the shoulders of giants: New approaches to numeracy. Washington, DC: National Academy Press. 232pp.

Nayak, B. K., & Hazra, A. (2011). How to choose the right statistical test.? Indian Journal Ophthalmology, 59(2), 85–86. https://doi.org/10.4103/0301-4738.77005

Paris J., Donnal J. S., & Leeb, S. B. (2014). The non-intrusive load monitor database Smart Grid, IEEE Transactions on Smart Grid 5(5), 2459-2467. https://doi.org/10.1109/ tsg.2014.2321582

Pettijohn, T. F., Naylor, P. D., & Piroch, J. F. (2013). Self-reported research skills changes and course outcomes in a senior research psychology course. American Journal of Educational Research, 1(3), 104-106. https://doi.org/10.12691/education-1-3-7

Phillip s-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. Journal of Decision Systems, 24(1), 87–102. https://doi.org/10.1080/12460125.2015.994333

Romero, C., & Ventura, S. (2012). Data mining in education. WIREs Data Mining and Knowledge Discovery, 3(1), 12–27. https://doi.org/10.1002/widm.1075

Safari, Y., Navazeshkhah, F., Azizi, M., Ziaei, M., & Sharafi, K. (2015). The Effective Factors on the Postgraduate Students' Interest and Participation in Performing Research Activities-Case Study: Kermanshah University of Medical Science. Future of Medical Education Journal 5(2), 3-9. https://dx.doi.org/10.22038/ fmej.2015.4454

Shah, T., Rabhi, F., & Ray, P. (2015). Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. Cluster Computing, 18(1), 351–367. https://doi.org/10.1007/s10586-014-0406-8

Shamoo, A. E., & Resnik, B. R. (2003). Responsible Conduct of Research. Oxford University Press. 360pp. https://doi.org/10.1093/acprof:oso/9780195368246.001.0001

Shepard, R.J. (2002). Ethics in exercise science research. Sports Medical Journal, 32 (3), 169-183. Retrieved from https://ori.hhs.gov/education/products/n_illinoisu/ datamanagement/datopic.html

Simonet, A., Fedak, G., & Ripeanu, M. (2015). Active Data: A Programming Model for Managing Big Data Life Cycle. [Research Report] RR-8062, 2012, pp.26.

Smith, J., Flowers, P., & Larkin, M. (2009). Interpretative phenomenological analysis: Theory, method and research: Qualitative Research in Psychology, 6(4), 346-347. https://doi.org/10.1080/14780880903340091

Thahir, M., Roza, Y., & Murni, A. (2018). Validity of learning website of kapita selekta Mathematics Course at UIN Suska Riau Students. Malikussaleh Journal of Mathematics Learning, 1(1), 19-25, https://doi.org/10.1088/1742-6596/1470/1/012092

UNESCO (2018). Technical and Vocational Education and Training: UNESCO Asia-Pacific in Graphic Detail Number 4, Retrieved from https://bangkok.unesco.org/ sites/ default/ files/assets/article/ Education/ files/module-5.pdf

Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). Focus group interviews in education and psychology. Newbury Park, CA: Sage.

Walliman, N. (2011). Research Methods: The Basics. Routledge, London. 205pp.

Wietse van der, L., Anouke, Bakx., Anje, Ros., Douwe,B., & Marc, V. (2012). Student teachers' development of a positive attitude towards research and research knowledge and skills. European Journal of Teacher Education 35(4), 401-419. https://doi.org/10.1080/02619768.2011.643401

Willison, J. & O'Regan K. (2007). Commonly known, commonly not known, totally unknown: a framework for students becoming researchers. High Education Research Development, 26(4), 393-409. https://doi.org/10.1080/07294360701658609

World Economic Forum (2021). "Data Science in the New Economy, Retrieved from http://www.weforum.org/docs/WEF_Data_Science_In_the_NewEconomy. pdf."

Zain, A. R. (2020). Challenges faced by students during conducting research: A review. International Journal of Multidisciplinary Research and Development, 7 (4), 16-19.

Zhang, F., Cao, J., Khan, S. U., Li, K. & Hwang, K. (2015). A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications. Future Generation Computer Systems, 43(44), 149-160. https://doi.org/10.1016/ j.future.2014.06.009

Zimbardi, K., & Myatt, P. (2014). Embedding undergraduate research experiences within the curriculum: A cross-disciplinary study of the key characteristics guiding implementation, Studies in Higher Education, 39(2), 233-250. https://doi.org/ 10.1080/03075079.2011.651448