Malaysian Journal of Mathematical Sciences 13(3): 419–437 (2019)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Effects of using RStudio on Statistics Performance of Malaysian Undergraduates

Salim, N. R.*[1], Gopal, K.[1], and Ayub, A. F. M.[1,3]

[1] *Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia*

[2] *Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia*

E-mail: nurraidah@upm.edu.my

* *Corresponding author*

Received: 31 August 2018

Accepted: 29 May 2019

ABSTRACT

This paper aims to study the effectiveness of using RStudio (an open source statistical package) in learning statistics among undergraduate students in Malaysia. A pre-post test quasi-experimental with control group design was employed. This study was conducted on 50 first year students from various fields of study taking the Statistics for Applied Sciences subject for selected topics from the syllabus. The students were equally assigned to the experimental and control groups respectively. Descriptive analysis showed that students performed better when using RStudio as compared to control group with higher means in post-test and post-delay-test scores. Analysis of Covariance (ANCOVA) was employed with the pre-test scores as covariate, group as the independent variable while post-test and post-delay-test scores as the dependent variables. ANCOVA indicated that there is a significant difference in the post-test scores between the groups as well as in the post-delay-test scores between the groups where students who has been exposed to and using RStudio achieved significantly better performance in pre-test and post-test. These findings imply that integration of a statistical package is effective in enhancing the learning process of statistics. In essence, statistical packages like RStudio would ease the learning process and help students to achieve better results in statistics.

Keywords: RStudio, statistics performance, analysis of covariance, statistical package.

1. Introduction

Ineffective teaching and learning process in introductory statistics courses among undergraduate students especially in the public universities in Malaysia has become an alarming and concerning issue lately. In particular, the learning of statistics among undergraduate students enrolled in non-mathematics/ statistics oriented programmes such as social sciences can be a difficult task. Instructors often find it hard to deliver their course contents effectively to the students. Some of them still regard the mathematical fundamentals of statistics alone are sufficient for novice students to have an essential understanding of statistics (Gould, 2010). On the one hand students do not comprehend statistics in the traditional learning ecosystem which is primarily based on the textbook approach and on the other hand, instructors are clueless in identifying the suitable channel for effective learning access and output as they tend to stick to the conventional instructional method. For instance, many challenges in statistics have been discussed by several researchers saying that students are facing anxiety in mathematics-based statistics, lack of content-specific focus in statistics classroom and many more (Strangfeld, 2013). Corresponded to that also, Horton and Hardin (2015) agreed that our graduates students nowadays do not acquire adequate abilities to be persuasive in their working environment particularly in handling real world data problems.

Most statistics instructors failed to realize that the current data-driven world has changed the role of statistics and data in the daily life eminently and frequently disdained the capability of technology. The availability of vast real life data requires students to understand them using computational tools for data management, exploratory analysis, visualisation, and modelling (Nolan and Lang, 2010; Horton, et al., 2015; Horton and Hardin, 2015; Ridgway, 2016). Diane Lambert of Google asserted students should learn how to utilise previous outcomes from existing data to drive statistical investigations and to decide on the choice of analysis; however, students with practically zero computational experience are not trained to do so (Wang et al., 2017). As a matter of fact, students need to perceive features of computation and data science which require statistical skills as early as possible (Horton et al., 2014). This is often important to develop deeper skills and to prepare themselves as prospective employees for the data-centric world. Hence, setting up forerunners in introductory statistics courses will ensure them to kick-start (Kim et al., 2018). Thus, the existing instructional method may not be apt anymore and needs to be reviewed. The persistence of such instructional method may result in reducing students' engagement in statistics classroom and eventually limiting the acquirement of proper statistical knowledge. Accordingly, in this era, learning statistics without using some computer programs or computational tools are hardly obtainable. Thus, shifting from the conventional approaches to widen the principles of effective teaching may require a suitable technology and using it with a proper Information and Communication Technology (ICT) tools and resources (Ertmer, 2010).

1.1 Role of Technology in Statistics Learning

ICT plays an important role as a tool to enhance the teaching and learning of statistics. The use of ICT with appropriate teaching method can give positive impact on students' cognitive and affective aspects of learning. Undeniably, the use of technology has become a part of everyday life especially in the education setting. Students can benefit more in their learning process if they are allowed to use technology in the classroom (Bloemsma, 2013). In particular, in teaching and learning statistics, students should use technology as a tool for expanding their understanding of the underlying mathematical concepts and promote active environment in learning statistics (Hassad, 2013). The impact of ICT itself has been investigated thoroughly among students to identify the implementation of laptop program by Stakkestad and Fladvad Stordal (2017) indicated that no significance effects among them. However, this program might affect and strengthen students' academic performance in a positive way like accessibility of current information, extra digital tools, and entertainment. In Al-Hariri and Al-Hattami (2017) affirmed that technology usage has a significant relationship between students and their achievements. Students' achievement had increased more than non-usage.

The importance of statistical tools in statistics classrooms were also discussed in Liu (2017). In fact, statistics academia was encouraged to grasp the use of technology in a wider perspective as the content and focus of the syllabus were constantly changing. Therefore, students' performance was measured more on their capability to use any relevant analysis and graphical tools for data explorations, data management, data interpretations and to come up with good conclusions. Furthermore, Nicholson et al. (2013) agreed that technology tools implemented for supporting statistical learning could bring exhilarating curriculum based on real-world scenario into the learning process and abled to trigger more on one-to-one interactions among the teachers and students. Parallel to that, students would have more space and freedom for feedback sessions, reflections during the classroom and revision at their own pace.

In line with this, the role of technology had been well endorsed in several instructional methods such as educational software, statistical software packages, spreadsheets, multimedia materials and others. The ability to perform among several tools like statistical packages, educational software, spreadsheets, applets/stand-alone applications, multimedia materials, graphing calculator and others slightly overlapped through the usage. Therefore, choosing the most appropriate technology is important to remain focus on the statistical content and not the type of tool, to achieve students learning goals. Previous researchers who sought the impact of technology towards achievement had conducted studies related to the use of technology. Meanwhile, getting the real statistics using technology into curriculum also being discussed deeply by Nicholson et al. (2013). Most of the studies comprised of using various types of mathematical and statistical software especially using open source software such as GeoGebra. For example, findings by Zamri (2017) on the effect of using GeoGebra software on students' conceptual and procedural knowledge indicated those who used GeoGebra to learn mathematics had higher mathematical conceptual and procedural knowledge compared to those who learnt mathematics through the conventional methods.

It was concluded GeoGebra software was capable of enhancing students' conceptual and procedural knowledge, which at the same time significantly improved students' achievement. Takači et al. (2015) also conducted a study to determine the effect of using GeoGebra on students' achievement in calculus. Their study indicated that by using GeoGebra software, it benefited students in their post-test scores compared to the control group. Similar study by Bhagat and Chang (2015) also revealed that GeoGebra gave a positive impact on students' achievement in learning geometry. Arbain and Shukor (2015) also showed students had positive perceptions towards learning and had better learning achievement using GeoGebra. A study by Leong (2013), conducted in Malaysia, compared two groups who used different approaches to teach graph functions. The results of this study showed the students who had better achievements were those who received mathematics and statistics instructions using Geometer Sketchpad (GSP), as compared to the control group whose lesson was conducted without using software technology. Besides, the use of GSP gave positive effect on students' attitude towards learning graph functions. In another study, Kesan and Caliscan (2013) determined the effects of GSP on students' achievement in the teaching geometry. The findings showed there was a meaningful difference in favour of the use of GSP.

1.2 Integration of Statistical Teaching Tools in The Classroom

Accordingly, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) curriculum framework for PreK-12 precisely advocated the use of technology to enhance students' understanding in statistical concepts and analysing data in teaching an introductory statistics course specifically for undergraduate students (Carver et al., 2016). In addition, according to Lawless and Pellegrino (2007); "technology can accelerate any routines abruptly in teaching or learning statistics, which makes it feasible to adopt current approaches to instruction and change the context of learning, instruction and assessment in statistics". Moreover, the changes in statistical routine would give an unambiguous impact content-wise in the introductory materials used. For example, using *ggplot* and *dplyr* packages in RStudio will allow students to grasp the concepts of probability theory and computations mainly for manipulating big data and developing interactive graphical output efficiently (Gunawan et al., 2018). Several researchers also asserted the utilization of contextual real life data sets accompanied by the appropriate use of technology i.e. practical data computing, could improve the traditional instruction method (Nolan and Lang, 2010; Carver et al., 2016; Liu, 2017). One of the viable ways to incorporate technology in statistics learning was by integrating statistical software in introductory statistics courses.

The first approach to integrate statistical software was statistical computing. In this approach, the course was taught as usual but the statistical software was primarily used as computational tools to replace by-hand calculations and manual plotting (Baglin, 2013). Statistical computing helped to speed up calculations (especially for large data sets) and to produce high quality plots. The second approach was known as computational statistics. Some fundamental changes to the contents of the course occurred in this approach. Essentially, the availability of statistical software as computation resource determined how the teaching was done (Baglin, 2013). Computational statistics was verily the core element of data science i.e. to answer questions using data and to convey the mitted results (Finzer, 2013). The common practice in most statistics courses was to combine aspects of both approaches (Gunawan et al., 2018). The proportion of aspects used from each approach depended on factors such as the aim of the course, resources of technology and the preference of instructors. In any case, it was inevitable to engage students in at least statistical computing

(Meletiou-Mavrotheris et al., 2007). However, the existing Malaysian academia emphasised statistical computing approach, in which undergraduate students were exposed on a minimal scale to statistical software (often SPSS) in their classroom learning. Through this practice, students focused on the interpretation of statistical results rather than stressing the mathematical derivations. Ideally, statistical software like SPSS was not designed as a teaching tool and it did not obscure statistical education as a whole (Cobb, 2007).

1.3 RStudio as A Teaching Tool

Among various statistical packages available, R is becoming immensely popular among the academia and industry community. R is an environment for statistical computation and graphics highly preferred by researchers for their statistical and data analysis tasks. Essentially it caters a programming language, high level graphics and visuals, interfaces to other programming languages and debugging facilities (Cook et al., 2007). Recent development of RStudio has made R seamlessly easier to access and used especially for novice students. RStudio is an integrated development environment (IDE) for R. It contains a console, syntax-highlighting editor that supports direct code execution, along with tools for plotting, history, debugging and workspace management (Verzani, 2011). Hereafter, the term RStudio will be used to denote both R and RStudio inclusively. RStudio is an excellent choice of statistical software as it is platform independent and more importantly, it is readily available at no cost (open source). On top of that, R is said to be relevant beyond classroom learning as it complies with any type and level of users; and powerful, flexible, and extensible with its own programming language (Schumacker, 2014). The major advantage of using RStudio is that it allows students to perform data analysis, computations and create plots by writing functions and scripts; and not by pointing and clicking or choosing an analysis from a dropdown menu, which is common in other software (Gunawan et al., 2018). Students get an interactive experience that encourages experimentation; exploration and play when working with RStudio i.e., it encompasses tools for thinking and learning with data (Schumacker, 2014). This process may sound intimidating to students who are new to programming, yet this is where students really get to learn and deeply understand the underlying statistical concepts as well as enabling them to think and reason with data (Horton and Kleinman, 2015). Introductory statistics courses can be a good platform to provide training in RStudio, as it is slowly becoming a norm for some instructors who used RStudio for their own statistics works to include the applications in their lectures too (Godfrey, 2013). Hence, RStudio may play a significant role in students' learning process if integrated into the fabric of introductory statistics courses (Kim, 2018; Gunawan et al., 2018).

One prominent effort to integrate RStudio in introductory statistics courses was initiated by project MOSAIC which consisted of a group of educators working to come up with innovative ways to introduce mathematics, statistics, computation, and modelling to college and university students (Pruim et al., 2014). The aim of project MOSAIC was to disseminate ideas and resources to enhance existing instructional methods and to develop a curricular and evaluation plan that helped the sharing and assessment of these ideas. Additionally, its goal was to create a far more extensive way to quantitative studies for the betterment of works related to science and technology. The focus of project MOSAIC was to combine various effective aspects of quantitative skills that would be beneficial for the professional lives of students in both science and social sciences (Pruim et al., 2014). In particular, for statistics, project MOSAIC focused on teaching introductory statistics courses that were highly integrated with computing elements in general and RStudio specifically. Accordingly, project MOSAIC created a package called mosaic which provided simplified and structured introduction to the essence of statistical concepts related to descriptive statistics, visualisation, modelling, and statistical inference (Pruim et al., 2016). The mosaic package helped novice students to get rid of unnecessary difficulties in using RStudio, as it was often regarded as unappealing and inaccessible to them. Students involved with project MOSAIC had indicated with a high level of confidence that it was viable to integrate computing into the syllabus early (Pruim et al., 2017).

In line with the above discussion, the objective of this study was to investigate the effectiveness of RStudio in statistics learning among undergraduate students in Malaysia, with a special focus on students in public universities who were enrolled in non-mathematics/ statistics oriented programmes.

2. Methodology

2.1 Study Design

This study employed a pre-test, post-test and post-delay-test quasi-experimental with control group design. This study was conducted on fifty randomly selected first year students from a broad spectrum of non-mathematics/ statistics oriented disciplines and specialisations taking an introductory statistics course in a Malaysian public university. The selected students were equally assigned to the experimental group (utilizing RStudio in their learning) and control group (did not utilize RStudio in their learning), n = 25 in each group.

Students' demographic information is provided in Table 1. There were seven male (14%) and forty-three female (86%) students. In terms of ethnicity, there were thirty-four Malays (68%), ten Chinese (20%), three Iban (6%) and one Indian, one Kadazan-Dusun (2%) and one Melanau (2%). The majority of the students were from the Faculty of Forestry (16%); followed by the Faculty of Educational Studies and the Faculty of Human Ecology (14% each); the Faculty of Agriculture (12%); the Faculty of Economics & Management (12%); the Faculty of Food Science & Technology (12%); the Faculty of Environmental Studies (12%) and the Faculty of Biotechnology and Biomolecular Sciences (8%).

Table 1: Demographic Information of Students in This Study (Frequency)

		Experimental Group	Control Group
Gender	Male	20	23
	Female	5	2
Ethnicity	Malay	18	16
	Chinese	6	4
	Indian	1	0
	Others	0	5
Faculty	Agriculture	3	3
	Forestry	4	4
	Economics & Management	3	3
	Educational Studies	3	4
	Food Science & Technology	3	3
	Human Ecology	4	3
	Biotechnology & Biomolecular Scineces	1	3
	Environmental Studies	4	2

2.2 Assessment

Students from both groups were required to sit for a pre-test, which was held before the lectures commenced on the first week of the semester. Following that, the experimental group underwent additional tutorial sessions in a computer laboratory for the new learning process that implemented RStudio. Prior to these RStudio tutorial classes, a special four-hour intensive session was conducted on the experimental group to introduce RStudio. This was to ensure the usage of RStudio would not limit the students' ability to enhance their understanding during the learning process. The special session and the RStudio tutorials were conducted by a statistics tutor who was trained in both statistics and RStudio. Six weeks after the pre-test, students sat for the post-test and two weeks later they sat for the post-delay-test which concluded the assessment.

Only five initial chapters from the course syllabus were chosen to assess the students. This is to avoid students to feel burdened with additional tasks aside from their usual learning routines. The chosen chapters were as follows:

i. *Describing Data with Graph* covering topics like frequency distribution table, relative frequency, histogram and ogive;

ii. *Describing Data with Numerical Measures* covering topics like descriptive statistics for grouped data and coefficient of variation;

iii. *Probability* covering topics like concept of probability, permutations and combinations;

iv. *Random Variables* covering topics like concept of random variables, discrete random variables and continuous random variables;

v. *Discrete Distributions* covering topics like Bernoulli distribution, Binomial distribution and Poisson distribution.

All three tests consisted of six questions (items) to evaluate students' understanding in solving the statistical problems arising from the selected chapters using RStudio. Subjective questions were provided with a maximum of three sub-questions. Example of question is provided in Figure 1.

Question 6

A population of 70 registered voters contains 40 in favor of 'Proposition 134' and 30 opposed. An opinion survey selects a random sample without replacement of 10 voters from this population.

a) what is the probability that there will be no one in favor of 'Proposition 134' in the sample?

b) What is the probability that there will be at least one person in favor?

Figure 1: Sample Question

The questions used in these tests were similar but the given problems were modified by changing the numerical figures and the problem context. As the contents of these tests were similar, several steps were taken to address threats to validity such as practice effect. These steps included constructing the questions according to the standard degree of difficulties of the syllabus, getting statistics experts to review and check the validity of the questions, and preparing the marking scheme based on the standard assessment procedure. Example of answer scheme is provided in Figure 2. All three test scores were recorded and statistical analyses were carried out to evaluate students' performance.

```
ANSWER 6

A)
> dhyper(x=0,m=40,n=30,k=10,log = FALSE)
[1] 7.573651e-05

B)
> phyper(q=0,m=40,n=30,k=10,lower.tail = FALSE)
[1] 0.9999243
OR
> 1-phyper(q=0,m=40,n=30,k=10,lower.tail = TRUE)
[1] 0.9999243
```

Figure 2: Sample Answer Scheme

2.3 Hypotheses and Statistical Analyses

In order to address the objective of this study, the following hypotheses were formulated:

H_{01}: There is no significant difference in the pre-test assessment scores (initial performance) among students in both groups;

H_{02}: There is no significant difference in the post-test assessment scores (medial performance) among students in both groups based on their pre-test assessment scores (initial performance);

H_{03}: There is no significant difference in the post-delay-test assessment scores (final performance) among students in both groups based on their pre-test assessment scores (initial performance);

H_{04}: There is no significant difference in the post-delay-test assessment scores (final performance) among students in both groups based on their post-test assessment scores (medial performance).

The formulated hypotheses were tested using the following statistical procedures:

i. Independent samples t-test on pre-test mean scores of both groups for H_{01};

ii. One way analysis of covariance (ANCOVA) with post-test scores as the dependent variable, grouping variable as the independent variable and pre-test scores as the covariate for H_{02};

iii. One way analysis of covariance (ANCOVA) with post-delay-test scores as the dependent variable, grouping variable as the independent variable and pre-test scores as the covariate for H_{03};

iv. One way analysis of covariance (ANCOVA) with post-delay-test scores as the dependent variable, grouping variable as the independent variable and post-test scores as the covariate for H_{04}.

3. Results

Table 2 depicts the independent sample t-test conducted on the pre-test mean scores of both groups to test the first hypothesis. It followed H_{01} was accepted since there was no statistically significant difference in the initial performance of students in both groups ($t = -.99, p > 0.05$). Thus, it could be inferred that all students had the same level of initial performance for statistics, prior to the integration of RStudio.

Table 2: Independent sample t-test on pre-test mean scores

Group	Mean	Standard Deviation	t	df	Significance
Experimental	38.00	10.39	-.99	48	.33
Control	33.00	22.92			

The mean scores of post-test assessment and corresponding standard deviation for both groups are displayed in Table 3. The experimental group students had a mean score of 67.30 ($SD = 20.60$) while the control group students had a mean score of 32.30($SD = 13.30$).

Table 3: Post-test Assessment Mean Scores and Standard Deviation

Group	Mean	Standard Deviation
Experimental	67.30	20.60
Control	32.30	13.30

Table 4 depicts the ANCOVA used to test H_{02}. It could be seen there was a statistically significant difference in students' post-test assessment scores between the groups based on their pre-test assessment scores $[F(1,47) = 51.81, p < .05]$ and thus, H_{02} was rejected. Further, Tukey post hoc test (pairwise comparison) revealed the experimental group students' post-test assessment scores were statistically significantly higher compared to those in the control group ($p < .05$) as depicted in Table 5.

Table 4: Test between subject effects of post-test scores based on pre-test scores

Source	Type III Sum of Squares	*df*	Mean Square	F	Significance
Corrected Model	17505.50	2	8752.75	33.90	.00
Intercept	12737.83	1	12737.83	49.37	.00
Pre-test	2227.98	1	2227.98	8.63	.01
Group	13375.13	1	13375.13	51.81	.00
Error	12134.51	47	258.18		
Total	153642.00	50			
Corrected Total	29640.00	49			

Table 5: Pairwise comparison for post-test scores

	Mean Difference	Standard Error	Significance
Experimental-Control	33.05	4.59	.00

As for the post-delay-test assessment, the mean score of the experimental group was $76.50 (SD = 14.00)$ and $52.80 (SD = 21.60)$ for the control group as shown in Table 6. Based on ANCOVA for testing H_{03} (Table 7), it could be observed there was a statistically significant difference in the post-delay-test scores between the groups based on their pre-test assessment scores $[F(1,47) = 19.80, p < .05]$, resulting in the rejection of H_{03}. Table 8 provides the Tukey post hoc test (pairwise comparison) indicating the experimental group students' post-delay-test assessment scores were statistically significantly higher compared to the students in the control group ($p < .05$).

Table 6: Post-delay-test Assessment Mean Scores and Standard Deviation

Group	Mean	Standard Deviation
Experimental	76.50	14.00
Control	52.80	21.60

Table 7: Test between subject effects of post-test scores based on pre-test scores

Source	Type III Sum of Squares	*df*	Mean Square	F	Significance
Corrected Model	8866.63	2	4433.32	14.84	.00
Intercept	26605.93	1	26605.93	89.06	.00
Pre-test	1833.65	1	1833.65	6.14	.02
Group	5919.09	1	5919.09	19.81	.00
Error	14040.59	47	298.74		
Total	231953.00	50			
Corrected Total	22907.22	49			

Table 8: Pairwise comparison for post-delay-test scores

	Mean Difference	Standard Error	Significance
Experimental-Control	21.98	4.94	.00

Table 9 details out the ANCOVA used to test H_{04}, in which it was revealed there was a statistically significant difference in the post-delay-test scores between the groups based on their post-test assessment scores $[F(1, 47) = 1.82, p < .05]$, hence H_{04} was rejected.

Table 9: Test between subject effects of post-delay-test scores based on post-test scores

Source	Type III Sum of Squares	*df*	Mean Square	F	Significance
Corrected Model	9487.12	2	4743.56	16.61	.00
Intercept	10082.01	1	10082.01	35.31	.00
Pre-test	2454.14	1	2454.14	8.60	.01
Group	520.36	1	520.36	1.82	.00
Error	13420.10	47	285.53		
Total	231953.00	50			
Corrected Total	22907.22	49			

4. Discussion and Conclusion

The objective of this paper was to evaluate the effectiveness of RStudio in statistics learning among undergraduate students in Malaysia, particularly in the public universities with a special focus on students enrolled in non-mathematics/statistics oriented programmes. Students from both groups were at the same level of initial performance for statistics, which ensured the conducted assessments to be meaningful to compare between both groups. Based on their pre-test assessment scores, it was observed that there were statistically significant differences in post-test and post-delay-test assessment scores between students who used RStudio and students who did not use RStudio in their learning. The mean scores for each assessment and pairwise comparison for both groups were in favour of students who used RStudio as their post-test and post-delay-test assessment scores were statistically significantly higher compared to students who did not use RStudio. Additionally, there was also statistically significant difference in post-delay-test assessment scores based on their post-test assessment between students in both groups. As for students' improvement in terms of their performance, it was revealed students who used RStudio had statistically significant improvement from pre-test to post-test assessments and as well as continued to improve from post-test to post-delay-test assessments.

Findings of this study implied students who used RStudio alongside their conventional teaching process performed significantly better compared to students taught using only the conventional teaching method. Furthermore, these students had also shown consistent improvement in terms of their performance in statistics. Hence, it could be affirmed the integration of RStudio was effective in enhancing the learning process of statistics and capable of sustaining students' performance. In essence, a learning strategy that blended the conventional learning process and integration of statistical packages like RStudio would enable students to achieve better results in statistics course by making it more attractive, interesting and engaging. Gunawan et al. (2018) also concluded, students preferred to explore statistics using RStudio more, as to compare with the traditional approach which only emphasized the theoretical content of Statistics during classroom. Many other literatures recently emphasized the use of R tools in statistical learning (Nicholson et al., 2013; Cobb, 2015; Hardin, 2016). Furthermore, Cobb (2015) mentioned, statistical packages provided close connection with the "experience" of a statistician, in which students would solve real world problems by working on real and substantive data analysis.

It is believed that the need for integrating statistical packages in statistics education will increase in the future as the world is becoming more data-centric and data-driven as well as in line with the availability of many data resources. The use of technology especially computer software also showed a positive impact towards teaching and learning. For example study by Zamri (2017) towards students conceptual and procedural knowledge, Arbain and Shukor (2017), Takači et al. (2015), Bhagat and Chang (2015), Leong (2013) and Kesan and Caliscan (2013) showed a significantly positive impact towards students achievement. Integration of technology in the form of statistical packages such as RStudio in learning of statistics will help to captivate students to be engaged and get motivated throughout their course completion (Kim et al., 2018).

Using RStudio provides a difference in students' learning approach as compared to the conventional textbook based instruction method. Many statistical inference techniques being a major part of introductory statistics courses were usually taught as disconnected topics in the traditional instruction method. This approach can be enormous to students; and hence, RStudio helps to unify statistical inference concepts by using just one built-in function that does it all.

This shows how RStudio can be handy for learning of statistics while still requiring students to think about the nature of the data and encourages them to perform exploratory data analysis. It is notable the computation approach offers a far more interesting and innovative ways of learning statistics as well as helps students to enhance their statistical thinking and reasoning. The use of statistical packages should be encouraged along with traditional instruction method and implemented with realistic, practical and related to daily life problem solving skills. RStudio also will help to shorten the learning curve in understanding statistics concept (Hardin, 2016). In this way, students will appreciate and apprehend to the overall learning process of statistics. Such initiatives are important because introductory statistics courses are perhaps the only source in equipping students with adequate statistical knowledge.

5. Recommendations and Conclusion

Statistics instructors are recommended to balance between traditional and computation components in their introductory statistics courses (Baglin, 2013). As in the case of RStudio which requires a fair amount of coding, instructors can use this opportunity to teach coding as a way to introduce or reinforce statistical concepts especially for concepts that are difficult to convey without computation like simulations (Navarro, 2015). Instructors who have previously taught their courses with minimal or zero integration of statistical software, may need to consider certain changes in their course.

For example, instructors can stop teaching students how to read statistical tables, since the use of tabulations of critical values of distributions is completely redundant now; and replace teaching of computational formulas with computation itself by focusing on how to lead to the understanding of the computed statistics instead of the computation process (Wang et al., 2017). At the same time, instructors should make use of packages to make other things possible like working with large data sets; creating better quality plots; teaching of simulation or resampling methods; perform quick computations; and finally concentrate more on concepts rather than calculations which may not be possible previously (Pruim et. al, 2017). Ideally, in any means of integration of statistical software in introductory statistics course, the sole purpose should be for students to learn statistics and not to learn a package or what it can do for them. To that end, the transformed syllabus is crucial in order to produce graduates that able to "think with data" and enriched with "deep analytical skills," as a preparation for them in the modernized personnel.

References

Al-Hariri, M. T. and Al-Hattami, A. A. (2017). Impact of students' use of technology on their learning achievements in physiology courses at the university of dammam. *Journal of Taibah University Medical Sciences*, 12(1):82–285.

Arbain, N. and Shukor, N. A. (2015). The effects of geogebra on students achievement. *Procedia-Social and Behavioral Sciences*, 172:208–214.

Baglin, J. (2013). Applying a theoretical model for explaining the development of technological skills in statistics education. *Technology Innovations in Statistics Education*, 7(2):1–7.

Bhagat, K. K. and Chang, C. (2015). Incorporating geogebra into geometry learning-a lesson from. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(1):77–86.

Bloemsma, M. S. (2013). *Student Engagement, 21st Century Skills, and How The iPad is.*. PhD thesis, Univeristy of California, Irvine.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G. H., Velleman, P., Witmer, J., and Wood, B. (2016). Guidelines for assessment and instruction in statistics education (gaise) college report. Retrieved from http://www.amstat.org/education/gaise.

Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1):Article 1.

Cobb, G. W. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, 69(4):266–282.

Cook, D., Swayne, D. F., and Buja, A. (2007). *Interactive and dynamic graphics for data analysis: with R and GGobi.* Springer Science & Business Media, New York, NY, 1st edition.

Ertmer, P. A. and Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of research on Technology in Education*, 42(3):255–284.

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, 7(2):1–9.

Godfrey, A. J. R. (2013). Statistical software from a blind person's perspective. *R Journal*, 5(1):73–79.

Gould, R. (2010). Statistics and the modern student. *International Statistical Review*, 78(2):297–315.

Gunawan, A., Cheong, M. L. F., and Poh, J. (2018). An essential applied statistical analysis course using rstudio with project-based learning for data science. *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 581–588.

Hardin, J. (2016). Dynamic data in the statistics classroom. arXiv preprint arXiv:1603.04912.

Hassad, R. A. (2013). Faculty attitude towards technology-assisted instruction for introductory statistics in the context of educational reform. *Technology Innovations in Statistics Education*, 7(2).

Horton, N. J., Baumer, B. S., and Wickham, H. (2014). Teaching precursors to data science in introductory and second courses in statistics. arXiv preprint arXiv:1401.3269.

Horton, N. J., Baumer, B. S., and Wickham, H. (2015). Setting the stage for data science: Integration of data management skills in introductory and second courses in statistics. *CHANCE*, 28(2):40–50.

Horton, N. J. and Hardin, J. S. (2015). Teaching the next generation of statistics students to "think with data": Special issue on statistics and the undergraduate curriculum. *The American Statistician*, 69(4):259–265.

Horton, N. J. and Kleinman, K. (2015). *Using R and RStudio for Data Management, Statistical Analysis, and Graphics.* CRC Press, New York, NY, 2nd edition.

Kesan, C. and Caliscan, S. (2013). The effect of learning geometry topics of 7th grade in primary education with dynamic geometer's sketchpad geometry software to success and retention. *Turkish Online Journal of Educational Technology-TOJET*, 12(1):131–138.

Kim, A. Y., Ismay, C., and Chunn, J. (2018). The fivethirtyeight r package:" tame data" principles for introductory statistics and data science courses. *Technology Innovations in Statistics Education*, 11(1):1–22.

Lawless, K. A. and Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research*, 77:575–614.

Leong, K. E. (2013). Impact of geometer's sketchpad on students achievement in graph functions. *The Malaysian Online Journal of Educational Technology*, 1(2):19–33.

Liu, Z. (2017). Teaching reform of business statistics in college and university. *Eurasia Journal of Mathematics Science and Technology Education*, 13(10):6901–6907.

Meletiou-Mavrotheris, M., Lee, C., and Fouladi, R. T. (2007). Introductory statistics, college student attitudes and knowledge–a qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology*, 38(1):65–83.

Navarro, D. (2015). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners. (Version 0.5).* University of Adelaide, Adelaide, Australia. R package version 0.5.

Nicholson, J., Ridgway, J., and McCusker, S. (2013). Getting real statistics into all curriculum subject areas: Can technology make this a reality?. *Technology Innovations in Statistics Education*, 7(2).

Nolan, D. and Lang, D. T. (2010). Computing in the statistics curricula. *The American Statistician*, 64(2):97–107.

Prium, R., Horton, N. J., and Kaplan, D. (2014). Start teaching with r. Project MOSAIC. Retrieved from http://mosaic-web.org/go/Master-Starting.pdf.

Prium, R., Kaplan, D., and Horton, N. J. (2016). mosaic: project mosaic statistics and mathematics teaching utilities. R package version 0.14.4.

Prium, R., Kaplan, D., and Horton, N. J. (2017). The mosaic package: Helping students to 'think with data' using r. *R Journal*, 9(1):77–102.

Ridgway, J. (2016). Implications of the data revolution for statistics education. *International Statistical Review*, 84(3):528–549.

Schumacker, R. E. (2014). *Learning Statistics using R.* Sage Publications, Los Angeles, CA, 2nd edition.

Stakkestad, S. V. and Fladyad Stordal, G. (2017). *The Effects of technology on students' academic performance rollout of individual laptops in norwegian upper secondary schools.* PhD thesis, Norwegian School of Economics, Norway.

Strangfeld, J. A. (2013). Promoting active learning: Student-led data gathering in undergraduate statistics. *Teaching Sociology*, 41(2):199–206.

Takaci, D., Stankov, G., and Milanovic, I. (2015). Efficiency of learning environment using geogebra when calculus contents are learned in collaborative groups. *Computers & Education*, 82:421–431.

Verzani, J. (2011). *Getting Started with RStudio.* O'Reilly Media, Inc., United States of America, 1st edition.

Wang, X., Rush, C., and Horton, N. J. (2017). Data visualization on day one: Bringing big ideas into intro stats early and often. *Technology Innovations in Statistics Education*, 10(1):1–18.

Zamri, S. N. A. S. (2017). Mathematics teachers' readiness toward information and communication technology application: A comparative study between malaysia and indonesia. *Advanced Science Letters*, 23(3):2033–2037.