Evaluation of data analytics workshop using RStudio amongst medical students in Pakistan

Manzar Abbas,[1] Hareem Rauf,[2] Javeria Bilal Qamar,[3] Syeda Ramlah Tul Sania,[4] Russell Seth Martins,[5] Zahra Hoodbhoy[6]

Abstract

A quasi-experimental study was conducted at the Aga Khan University, Karachi, Pakistan, to evaluate the outcomes of a series of workshops on 25 medical students' statistical knowledge and acceptance of RStudio. The knowledge in each of the five sessions was assessed using pre- and post-knowledge-based quizzes. In addition, the Student's Attitude Towards Statistics (SATS-36) and the Technology Acceptance Model were used. Data analysis on RStudio revealed a statistically significant improvement in knowledge in all five sessions ($p<0.05$). SATS-36 showed statistically significant improvement in Cognitive Competence ($p<0.001$). RStudio had commendable acceptance with relatively high scores of Attitudes (behavioural intention, median = 6.00 [5.20-7.00]) and Utility (perceived usefulness, median = 5.20 [4.10-6.20]). In conclusion, medical students had improved statistical knowledge and acceptance towards the novel statistical tool. Hence, further studies must evaluate the effectiveness of RStudio when integrated as part of the medical curriculum.

Keywords: RStudio; Technology acceptance; Peer learning; Statistics.

DOI: 10.47391/JPMA.6450

Submission completion date: 02-02-2022
Acceptance date: 11-08-2022

Introduction

With increasing advancements in technology and healthcare, there has been great emphasis on doctors' learning statistics and their beneficial effects on major research undertakings, such as randomised controlled trials (RCTs).[1] Swift et al highlighted that 90% of doctors considered statistical knowledge to be useful in various areas of clinical practice, including interpreting clinical guidelines, evidence-based summaries, explaining risk factors to patients, interpreting results of screening tests, reading research publications for general professional interest, and using research publications to explore non-standard treatment and management options.[2]

R, a free software, commonly used with RStudio is emerging as a new tool for statistical analysis, providing community support and possibilities for data visualisation.[3]

A study conducted by da Silva and Moura explored how a course teaching RStudio helped improve students' attitudes towards R as a statistical tool.[4]

There have been no such studies in Pakistan, so we aimed to assess the change in statistical knowledge and attitudes of medical students in Pakistan after attending a series of workshops teaching RStudio for statistical analysis.

Methods and Results

Five workshops were conducted for 25 medical students in July-August 2021 using a quasi-experimental study design at the Aga Khan University in Karachi, Pakistan. The learning objectives of each session were developed under the supervision of an expert researcher and a statistician at the university. The first three sessions included a brief introduction to statistics and an introduction to R and RStudio. The final two sessions focused on statistical analysis and data visualisation.

A pre- and post- quiz, consisting of 10 knowledge-based questions, was administered for each session. This methodology and questions were adopted from an assessment guide proposed by Garfield[5] and Hubbard[6] and R's official website.

Students' Attitudes Towards Statistics questionnaire (SATS-36) was administered before the first and after the last session.[7] It consists of 36 component questions that make up six dimensions: Affect, Cognitive Competence, Value, Difficulty, Interest, and Effort. At the end of the workshop, the participants filled out the Technology Acceptance Model (TAM) questionnaire[4] consisting of 15 component questions that make up three dimensions: Attitudes, Utility, and Facility.

[1,4]5th Year MBBS Student, [2]4th Year MBBS Student, [3]3rd Year MBBS Student, [6]Department of Pediatrics and Child Health, [5]Aga Khan University, Karachi, Pakistan.

Correspondence: Zahra Hoodbhoy. Email: zahra.hoodbhoy@aku.edu
ORCID ID. 0000-0002-0439-8293

Table-1: SATS-36 Results Summary.

SATS	Median [IQR] Pre-Test	Median [IQR] Post-Test	p-value	Effect Size
Affect	5.28 [4.40-5.87]	5.8 [4.78-6.38]	0.0741	0.255
Cognitive Competence	3.24 [2.54-4.11]	5.82 [5.28-6.45]	<0.001	2.541
Value	5.69 [5.45-5.98]	5.7 [5.37-6.11]	0.236	0.19
Difficulty	3.49 [2.73-4.02]	3.73 [3.04-4.51]	0.459	0.213
Interest	6 [5.25-6.87]	6 [5.50-7.00]	0.752	0
Effort	6 [5.00-7.00]	4.25 [3.00-5.12]	0.006	1.178

The analysis was done on R, R Core Team (2020).[8] Pre- and post-quiz scores and SATS-36 scores were compared using Wilcoxon's test. The effect size for intra-group pre- and post-scores was calculated using Cohen's d values, and the effect size for inter-group comparison was calculated using Hedges' g value (values greater than 0.80 and 1.30 were considered large, and very large, respectively).

Results from TAM were analysed descriptively. Scores were divided into two clusters (high and low), using k-means clustering, an unsupervised machine learning technique. The SATS-36 scores were stratified and analysed using these clusters. TAM results of each dimension for males and females were compared using the Mann-Whitney U test.

Out of 25 participants, 16 (64%) were males, and 9 (36%) were females. Of these, 9 (36%) participants were first-year medical students, and 13 (52%) were second-year medical students.

All five sessions had a statistically significant effect on students' knowledge, with the greatest median score increase in sessions 2, 3, and 4 (p<0.001). Comparing the differences in pre- and post-scores of males and females did not reveal any statistically significant differences.

Students' feelings towards statistics (Affect) did not show significant improvement in the pre- and post-scores. There was a statistically significant improvement in the self-perception of their intellectual knowledge and skills for statistics (Cognitive Competence). It was noted that the students did not put in the amount of work in learning statistics as they initially intended to, as seen in decreased Effort scores. The rest of the attitudes towards statistics (Value, Difficulty, and Interest) showed no changes, as shown in Table-1.

For TAM, participants had an overall positive Attitude toward RStudio (median = 6.00 [5.20-7.00]), and their perception of its usefulness (median = 5.20 [4.10-6.20]) was relatively higher than their perceived ease of use of the software (median = 4.60 [3.30-5.80]). There were no significant differences between the mean scores of male and female students in any three dimensions.

Participants who had a high acceptance of RStudio had a greater improvement in SATS-36 dimensions. Clusters with high acceptance had a statistically significant improvement in the median scores of Affect and Cognitive Competence on SATS-36 (Table-2).

Discussion and Conclusion

Our findings suggest that the RStudio workshop course proved to be effective for students as they were more confident about their knowledge and skills in statistics after the series. The participants developed positive feelings toward statistics and found statistics easier after the course as demonstrated by an improvement in the median score

Table-2: SATS-36 Results based on Degree of Acceptance of RStudio.

	High (n = 12)				Low (n = 7)			
SATS	Median [IQR] Pre-Test	Median [IQR] Post-Test	p-value	Effect Size	Median [IQR] Pre-Test	Median [IQR] Post-Test	p-value	Effect Size
Affect	5.44 [3.41-6.22]	6.04 [3.94-5.41]	0.042	0.668	5.16 [3.41-6.22]	4.06 [3.94-5.41]	0.834	0.117
Cognitive Competence	2.67 [2.91-3.88]	6.1 [4.21-5.64]	<0.001	3.937	3.42 [2.91-3.88]	5 [4.21-5.64]	0.078	1.448
Value	5.85 [5.21-5.69]	5.75 [5.36-6.11]	0.518	0.157	5.45 [5.21-5.69]	5.64 [5.36-6.11]	0.281	0.26
Difficulty	3.42 [2.73-3.98]	3.82 [3.21-4.39]	0.824	0.138	3.73 [2.73-3.98]	3.65 [3.21-4.39]	0.578	0.416
Interest	6.25 [5.37-6.50]	6.87 [4.00-6.00]	0.622	0.316	5.75 [5.37-6.50]	5.75 [4.00-6.00]	0.276	0.525
Effort	6.5 [5.00-6.50]	4.37 [2.62-5.12]	0.029	1.224	5.5 [5.00-6.50]	4.25 [2.62-5.12]	0.156	1.05

for Cognitive Competence and Affect on the SATS-36 questionnaire. This is similar to findings from a meta-analysis where students showed the most significant improvement in Cognitive Competence, which directly correlated with the student's overall statistics achievement.[9] There was an increase in the positive feelings toward statistics and attitudes about the Difficulty of statistics.[10]

The students did not put in the amount of work in learning statistics as they initially intended to, as seen in decreased Effort scores. However, the drop in the Effort scores post-test could be because of conflicts between time and other academic responsibilities. Few existing studies report an association with the age of the participants, which could be one of the reasons for these results.[9]

No significant changes were noted in the median scores for Value and Interest, which is in line with other similar studies done previously.[9]

Participants showed high acceptability for RStudio. Da Silva et al evaluated the technology acceptance of RStudio amongst medical students who reported high scores on the Attitude dimension, indicating that participants believed that RStudio was a helpful software that made statistics classes more interesting. The scores on the Utility dimension were also high, indicating that students considered RStudio an essential part of improving skills in statistics.[4]

Participants who had high acceptance of RStudio had significantly improved Affect and Cognitive Competence towards statistics. This was also reported by da Silva et al. In contrast, moderate scorers had statistically significant differences only in the Cognitive Competence dimension.[4] One possible explanation for this could be that the acceptance of technology could impact the students' attitudes towards and efforts invested in learning statistics.[4]

In summary, students had improved statistical knowledge, skills, and acceptance of RStudio. Given its flexibility for analysis, compatibility with multiple operating systems, and a wide variety of functions for statistical analysis and machine learning, further studies must evaluate the effectiveness of RStudio when integrated as part of the medical curriculum.

The limitations of the study include its small sample size; only students from a single institute were included. Importantly, our TAM was administered after the workshop series, rendering the relationship between acceptance and improvement in statistical skills unclear. Moreover, there was no control group against which our cohort's TAM results could be compared.

Our study is the first from Pakistan exploring the utility of RStudio for teaching statistical analysis to medical students and setting the groundwork for future large-scale studies assessing the impact of integrating similar workshops as part of the medical school curriculum.

Acknowledgement: We appreciate the efforts of Javerya Hassan, Hajra Arshad, Fatima Zahra, and Dr. Waqasuddin Khan in helping us in the study. We thank Dr. Schau for allowing us to use SATS-36 questionnaires.

Disclaimer: An abstract of this paper was presented at the 5th Educational Research Day, held at the Aga Khan University, Karachi, Pakistan in October 2021.

Conflict of Interest: None.

Funding Disclosure: None.

References

1. Altman DG, Bland JM. Improving Doctors' Understanding of Statistics.[Online] 1991 [Cited 2021 Aug 28]. Available from: URL: https://www.jstor.org/stable/2983040?origin=crossref
2. Swift L, Miles S, Price GM, Shepstone L, Leinster SJ. Do doctors need statistics? Doctors' use of and attitudes to probability and statistics. Stat Med. 2009; 28:1969–81. doi: 10.1002/sim.3608.
3. R vs SPSS Which One is The Best Statistical Language [Internet]. Stat Analytica.[Online] [Cited 2021 September 06]. Available from: URL: https://statanalytica.com/blog/r-vs-spss/
4. da Silva HA, Moura AS. Teaching Introductory Statistical Classes in Medical Schools Using RStudio and R Statistical Language: Evaluating Technology Acceptance and Change in Attitude Toward Statistics. J Stat Educ. 2020; 28:212–9.doi: https://doi.org/10.1080/10691898.2020.1773354
5. Garfield BJ. Beyond Testing and Grading: Using Assessment To Improve Student Learning. [Online] [Cited 2021 October 12]. Available from: URL: https://www.tandfonline.com/doi/abs/10.1080/10691898.1994.11910462
6. Hubbard R. Assessment and the Process of Learning Statistics. [Online] [Cited 2021 September 26]. Available from: URL: https://www.tandfonline.com/doi/abs/10.1080/10691898.1997.11910522
7. Schau C. Survey of Attitudes Toward Statistics (SATS-36).[Online] [Cited 2022 July 22]. Available from: URL: http://evaluationandstatistics.com
8. Bunn A, Korpela M. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Online] [Cited 2014 April 25].Available from: URL: https://www.r-project.org/
9. Milic NM, Masic S, Milin-Lazovic J, Trajkovic G, Bukumiric Z, Savic M, et al. The importance of medical students' attitudes regarding cognitive competence for teaching applied statistics: Multi-site study and meta-analysis. PLoS One. 2016; 11:e0164439. doi: 10.1371/journal.pone.0164439.
10. Schau C, Emmioğlu E. DoIntroductory Statistics Coursesinthe United States Improve Students' Attitudes?[Online] 2012 [Cited 2021 August 28]. Available from: URL: http://www.stat.auckland.ac.nz/serj