Original Research Paper

Statistical Demonstration of Superiority of High-Frequency Short-Tests in Assessment When Teaching Statistics to Enhance Higher Student Pass Rates

**Solly Matshonisa Seeletse, Tshepo Matsose,
Gezani Richman Miyambu and Eunice Lebogang Sesale**

*Department of Statistics and Operations Research,
Sefako Makgatho Health Sciences University, PO Box 107, MEDUNSA, 0204, South Africa*

Abstract: When facilitating learning, quick and small assessments can assist to detect areas that require further reinforcement and also identify students who need extra help in understanding the concepts taught. This paper demonstrates with Statistics, a difficult subject to pass at Higher Education Institution (HEI) level, that when these tests are administered at high frequencies and in large numbers, students are boosted to perform better, seemingly because there is no chance to pause learning. In addition to the higher performance shown, the students' marks were more stable when checked using the Coefficient of Variation (CV) measure. This was an indication that in the case of many assessment exercises, the higher student performances obtained were also more reliable and had a better chance to be replicated when the exercise was repeated.

Keywords: Assessment, Descriptive Statistics, High-Frequency Short-Tests, Statistical Tests

Introduction

Planning and managing every aspect of human life can be made easy with Statistics, which is a vital subject in HEIs (Umameh, 2011). Therefore, Statistics literacy is important for every person who cannot reach expertise in the subject. On the other hand, ideally, Statistics expertise is important for use in various sectors and industries such as business and government departments, among others. Historically though, this subject has high failure rates in HEIs. Makgato and Mji (2006) indicate that in South African HEIs, students fail Statistics at a very high rate. Various interventions to improve pass rates were tried in the past, including researches to investigate reasons for the high failure rates (Attwood, 2014). However, recently, several uses of assessment methods have been attempted with the aim to improve the pass rates in this subject.

This paper applies tests of hypotheses and statistical methods to compare the approach of giving many short tests after chunks of learning various Statistics component concepts relative to the approach of using few long tests in assessing students in Statistics. It fundamentally demonstrates that the former approach in assessing leads to better student performance in the Statistics subject.

Considerations When Teaching Statistics

Maths Subject

Statistics is a mathematics requiring learning of concepts through constant practice to enhance student performance. Various researchers (Benson, 2011; Manoah *et al*., 2011; Miheso, 2012; Makgato and Mji, 2006) have identified factors believed to cause poor performance as facilitators not using student-centred tactics, lack of experiments and practical modelling activities and lack of professional exposures to articulate issues related to the subjects. Eshiwani (2001) points out that poor performance is due to poor teaching methods, inadequate exercises and an acute shortage of books. Furthermore, reinforcing teaching is not included on the students' experiences. This, in some cases, consistently delays the pace to cover the full syllabus and thus leading to poor student performance.

Syllabus Coverage

Tswani (2009) counsels that rigorous teaching and learning principles promotes an environment of motivation to achieve. Buschang *et al*. (2012) highlight also that learning reinforcement was lacking or inadequate and students receive inadequate feedback on their work. As an intervention towards reducing the

failure rate, there could be use of teaching variations (Manapure, 2011; Gasca, 2011; Jennison and Beswick, 2010). The syllabus should also be fully covered (Slattery and Carlson, 2005).

Assessment

Kousha and Thelwall (2008) point out that understanding is not easy to measure until some assessment has taken place. Assessment verifies the level of knowledge and skills acquired in the subject and in measurable terms (Nelson and Dawson, 2014). According to Carless (2015), assessment can be used to diagnose areas requiring reinforcement and students still needing additional assistance. However, attempts to cover the full syllabus are usually a disturbance of concept understanding because in some cases, rushing to complete the syllabus may prevent understanding the content.

Learning Reinforcement

Reinforcing learning is effective in consolidating learning (Newcomer, 2009). Every class is different. Thus, different reinforcements may be used to motivate different students. In HEI, rein forcers should be straightforward and enticing to learn, such as blended learning methods, including interactive methods using many computing methods and tools to enhance reinforcement.

Research Method

Study Design

Experimental research suited this study, which was qualitative design necessitated by the interest to uncover details of the experimental exercise. The experimental group were students who were subjected to weekly short tests named High-Frequency Short-Test (HFST) assessment.

Sampling Design

The sample consisted of 1587 Statistics students in five (5) South African universities. The distribution of the sample was eight (8) student groups doing first to third years of the Bachelor of Technology (B Tech), Bachelor of Science (BSc), Bachelor of Arts in Personnel Management (BAPM) and Bachelor of Commerce (B Com) degrees. The experimental groups consisted of four student groups who were subjected to HFST assessments. The control groups consisted of four groups whose assessments were conducted routinely with two tests prescribed by the HEIs concerned. The experimental groups were selected deliberately as part of the experiment while the control groups were chosen because they were taking place parallel to the experimental groups.

Sample

One experimental group involved 179 first-year B Com students in an Eastern Cape HEI in 2008 with its counterpart control being 153 first-years of BAPM in the same university in 2007. The second experimental group consisted of 107 second-years of BSc from a HEI in the Vaal region of Gauteng Province in 2011. Its counter control group consisted of 54 third-years of BSc from a HEI in a Gauteng HEI. The third experimental consisted of 79 first-years of BSc in a Limpopo HEI with a control of 84 second-years of BSc in another Limpopo HEI. The fourth experimental group consisted of 423 first-years of B Tech of a HEI in North-West Province with a counter control group of 508 first-years of a B Com in another North-West HEI in year 2015. There was also a qualitative section where the respondents were facilitators in the modules of the experimental groups.

Research Instruments

Primary data collection tools were the mark sheets for experimental and control groups on final marks constructed from weekly assessments and compounding examinations. There was a question guide used for standardisation of responses, which was given to the students who participated in the experimental groups.

Data Collection

The numeric data were student marks recorded on mark sheets of tests and examination marks of the Statistics modules. Actual research data were final course numeric results obtained from sums of weighted marks of formative and summative assessments. The qualitative responses were collected from experimental groups by describing their experiences of being assessed by HFSTs.

Data Analysis

The quantitative section used descriptive statistics, Analysis of Variance (ANOVA), statistical tests of equality of means and statistical independence. Graphs were also utilized to illustrate and reveal some more information for the study. The qualitative section analysed responses on 'what the HFST mode of assessment was viewed and experienced.

Results

These refer to the section of the study which requires numeric responses.

From Table 1, all the marks in the Experimental Group (EG) exceed 88% while all the Control Group (CG) ones are below 82%. Also, all the maximum marks of the EG exceed all of CG and all the mean marks of the experimental groups exceed all of the control groups.

In Table 2, the Coefficients of Variation (CVs) of Experimental Group (EG) members are much lower than those of the Control Group (CG) ones. This indicates lower stability of the values in the experimental groups, where the CV value gives the precision of any measuring instrument or sampling procedure used (Armitage 2005; Kleijnen and Sargent 2000) and is defined by:

$$CV = \frac{Standard\ deviation}{Mean} \qquad (1)$$

This signal is supported by the standard errors of the experimental group from Table 1, which are all approximately 1 while for the control group they are all higher, closer to 1.5. This indicates that the experimental group approach is more consistent.

Further analyses follow to provide further insight into the investigation of the effect of using the small test high frequency (STHF)mode in enhancing performance in the Statistics subject. First, the bar charts displaying the comparisons for the various performances are provided below. The bar chart is based on Table 3, which was extracted from Table 1.

In reiterating the findings from Table 1, Table 3 shows that minimum marks of EG are mostly around 50% while those of CG are close to 40%; the EG maxima are all around 90% compared to EC ones which are all around 80%; the EG means being close to 60% while CG ones are close to 45%; and the EG passes are close to 90% compared to the EC ones which are below 80% to as low as below 70%. Analysis of variance (ANOVA) will be used to determine these further. ANOVA is discussed next and also its requirements are presented below.

Analysis of Variance

ANOVA entails a collection of statistical models for analyzing the differences in group means (Gelman 2008). It therefore generalizes the t-test to more than two groups. ANOVA is conducted under some assumptions. These assumptions are that random sampling is used on the populations being compared, observations should be independent, variables need to be normally distributed, there should be homogeneous variances and there needs to be sample size of at least 20 per cell (Bailey 2008; Van Belle, 2008).

On the study sample data fulfillment of the ANOVA requirements, random sampling is relieved by that a census (i.e., all members in the population of study) of all the groups studied was sampled. Therefore, the groups participating are fully represented. From Table 1, the groups of qualifications in the study are independent. Also, column n of Table 1 shows that all samples are over 20. Since each group of qualification consists of identical features, the clusters are uniform and are thus homogeneous in variances.

The remaining condition to be verified is normality. Fischer (2011) points out that the law of large numbers and the central limit theorem lead large samples convergence to approximate normality. Gelman (2005; 2008) derived that minor deviations from normality provide similar results when conducted under normality. Further, Fay and Proschan (2010) point out that the application of normal distribution of parametric methods is equally valid on categorical data of large samples sizes. Since using ANOVA requires a minimum of 20 observations for each category and the smallest size is 54, the sample sizes are large. Thus the categories of groups and qualifications in Table 1 exhibit the ANOVA features. They are also consistent with the methods of Kleijnen and Sargent (2000) and Reed *et al.* (2002) on the use of ANOVA. Therefore, using ANOVA is fully justified in the analyses of this paper.

From Table 4, the sums, averages and variances of the minimum marks, maximum marks, mean marks and pass percentages differ according to groups. These are indications of being different. However, the ANOVA differ according to groups to determine the statistical significance of these differences. The null hypothesis being tested in ANOVA is that the mean values of the groups are equal.

The p-value in Table 5 is almost 0 and the value of the F statistic far exceeds the critical F value. The two facts concur that there is no statistical evidence at the 5% level of significance that the mean values are identical.

Further Comparisons

The entire EG were 788 HFST students while the CG were 799 students who experienced normal assessments. The mean from the groups where various subgroups are involved is defined by:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \qquad (2)$$

Where:
x_i = Value of item being averaged, $i = 1, 2, \ldots, n;$ and
f_i = Frequency corresponding to x_i

The averages of minimum marks, maximum marks, mean, standard deviation, standard error and pass percentage rate are calculated and presented in the next table.

The experimental group's average minimum mark, average maximum mark, average of the mean marks and the average pass percentages are all higher than the corresponding statistics of the control groups. This fact is demonstration of superiority of performance of the EGs.

Table 1. Statistical results

Group/ var/s*	Qualif	n	Level	Prov	Min	Max	Mean	Stddev	Std err	Pass %
EG	B Com	179	1	EC	44	93	58.79	11.6213	1.0321	93.85
	BSc	107	2	Gau	51	89	61.07	12.3215	0.9943	89.72
	BSc	79	1	Lim	49	92	56.22	11.2345	1.0094	88.61
	B Tech	423	1	NW	42	91	53.25	10.4491	0.8934	92.36
CG	BAPM	153	1	EC	39	79	43.94	37.3604	1.7723	65.46
	BSc	54	3	Lim	47	81	46.41	40.1429	1.6908	81.48
	BSc	84	2	Lim	38	78	49.68	46.3217	1.4409	76.19
	B Com	508	1	NW	34	82	48.32	41.3096	1.9428	69.68

Table 2. Coefficient of variation

	Experimental group				Control group			
Course	BCom1	BSc2	BSc1	BTech1	MAPM 1	BSc3	BSc 2	BCom1
CV	0.1977	0.2018	0.1998	0.1962	0.8503	0.865	0.9324	0.8549

Table 3. Group performances

Group/variables*	Qualif	Performance			
		Min mark	Max mark	Mean	Pass (%)
Exp. group	B Com	44	93	58.79	93.85
	BSc	51	89	61.07	89.72
	BSc	49	92	56.22	88.61
	B Tech	42	91	53.25	92.36
Control group	BAPM	39	79	43.94	65.46
	BSc	47	81	46.41	81.48
	BSc	38	78	49.68	76.19
	B Com	34	82	48.32	69.68

Table 4. Comparison of summary values

Groups	Count	Sum	Average	Variance
Min mark	8	344.00	43.00000	34.2857
Max mark	8	685.00	85.62500	38.8393
Mean	8	417.68	52.21000	37.5034
Pass %	8	657.35	82.16875	115.7422

Table 5. ANOVA

Source of variation	SS	df	MS	F	p-value	F crit
Between groups	10923.88	3	3641.29	64.34	1.12E-12	2.946685
Within groups	1584.594	28	56.59			
Total	12508.47	31				

Table 6. Averages of values of groups

	Min	Max	Mean	Std. dev	Std. Err	Pass (%)
EG	44.4	91.3	55.87	11.05	0.95	91.96
CG	36.26	80.94	47.50	41.00	1.84	70.35

Table 6 shows that the average standard deviation and average standard error of the EG are lower than the corresponding ones of the CG. Initially, this shows stability of the EG values, which indicates more reliability of the values obtained. In using the CV in equation (1), these are in Table 7.

Table 7 shows that the CV corresponding to EG is much smaller than the one for CG. Therefore, the EGs producehigher student performances in Statistics than the CGs. These higher performances are also more stable when compared with the CG ones.

Table 7. Averages of CVs

	EG	CG
CV	0.0140	0.0513

Figure 1 below shows that the performances of students on HFST were performing better than the group where fewer tests were given.

Test of Independence

The chi-square test of independence is used next to determine if performances depend on the group (Wackerly *et al.*, 2008) the null hypothesis being tested is.
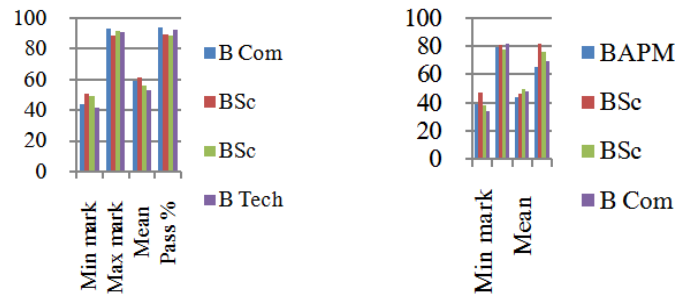
Fig 1. Visual displays of comparative performances

H_0 : Performances using HFST and student group option are independent
H_a : Performances from STHF and student group option are not independent

Let o_i denote the ith observed frequencies and e_i the corresponding expected frequencies in a contingency table (e.g. frequency table). The value of the test statistic (Bless *et al.*, 2006) is given by the formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad (3)$$

The table of marks to be used and the one on expected values based on the hypothesis of independence, follows:

Based on these analyses, equation (1) gives $\chi^2 = 5.3067$. The degrees of freedom (d.f.) of the chi-square statistics is $(c-1)(r-1)$ when the contingency table of a matrix with c being the number of columns and rows (Tabachnick and Fidell, 2007). Then the value of the d.f. = (c-1)(r-1) = (4-1)(8-1) = 24. Thus, from Kutner *et al.* (2005), the critical value at 5% level of significance with 16 d.f. is $\chi^2_{0.05} = 13.8484$. The test statistic does not exceed the critical value and therefore the null hypothesis cannot be rejected. This result confirms that there is enough statistical evidence at the 5% level of significance that performance depends on the group being assessed.

Challenges in Implementing HFSTs

The mode with HFST patterns is labour intensive. It requires many student support staff to assist in marking and tutorial sessions for revisions. This requires teamwork, which may be difficult in some instances. For this study, student support involved tutors who were students enrolled for postgraduate degrees in the subject. The advantage of using these tutors was that they were paid per claim of the work done. Their reappointment depended on performance. Tutors were thus encouraged to work hard and justify their payment and their reappointments. The other challenge was that these HFST requires a high level of documentation. Hence, in order to ensure a common and consistent understanding for facilitation, the lecturer wrote amply to communicate messages, feedback, updates on exercises and clarifications, among other assurances.

Benefits to Role Players

The department trained the tutors to become academics and statistics practitioners. The HFST mode was therefore beneficial for giving experiences and training to the Statistics students. The recruitment of the departments was also somewhat biased towards the best students who emerged from the HEIs. Thus, the HFST model offsets understaffing as tutors replace lecturers on some minor academic activities.

Discussion

The HFST assessment mode is absolutely beneficial in facilitating the Statistics subject. More students pass and with high marks. The method also enhances completion of the syllabus. It also increases the understanding of the subject. The results of this method are also more stable. The HFST assessment mode is labour intensive and requires more staff to be used.

The HFST feedback gives students an opportunity of success by pointing at their needs at an early stage of their Statistics learning. Another HFST benefit is teamwork, mainly because it is required when executing the HFST assessment mode. The team members, who are usually the tutors, are being groomed and therefore gain academic experiences. The benefit is therefore giving the tutors some academic experience. It thus makes them more marketable when they apply for employment posts of statisticians, especially in academia.

Recommendations

Based on the study finding findings, recommendations are made to the lecturers facilitating Statistics with a view to improve performance of students enrolled in the subject. The recommendations are that:

- Students of Statistics should receive many tests on a regular short intervals basis to ensure no pausing in learning

- The HFSTs should be incorporated in the study materials as part of the lectures prepared
- Departments should involve tutors who could assist with marking of the HFSTs with feedback
- Feedback should be prepared timeously, in detailed form and reinforcing material be added; and
- The facilitator should exercise regular written communication with student support role players

Conclusion

Forms of assessment can be used creatively to reinforce learning of difficult science subjects. When Statistics students are assessed regularly using small assessment tests without a break in their learning, they tend to have a larger chance to perform better. The HFST assessment mode encourages students to learn the concepts early in their learning and enables facilitators to identify problem areas and at-risk students. Also, this mode enables the completion of the syllabus in Statistics. The mode also enhances learning of the concepts. At-risk students are easy to identify and reinforcement can occur. Understanding shows to be higher with the HFST assessment mode. The student performance in the Statistics subject is also higher with this assessment method.

Acknowledgement

Funding Information

Authors' Contributions

Solly Matshonisa Seeletse: Advised on various aspects of this paper and identified AJAS for this paper.

Tshepo Matsose: Tested the data for outliers based on her studies on 'Biases from poor data analyses'.

Gezani Richman Miyambu: Verified that the methods of this study provided robust analyses.

Eunice Lebogang Sesale: Suggested statistical tests on previous work done and the compilation of the work as a journal manuscript.

Ethics

The agreements have been observed and as agreed before using the data, the identities of participants have been concealed. There will simply be no ethical issues after publication of this manuscript.

References

Armitage, C.J., 2005. Can the theory of planned behavior predict the maintenance of physical activity? Health Psychol., 24: 235-245.

Attwood, T., 2014. Why are some students so poor at maths?

Bailey, R.A., 2008. Design of Comparative Experiments. 1st Edn., Cambridge University Press, London.

Benson, S.K., 2011. Factors contributing to poor performance in mathematics at KCPE in Public primary schools in Isiolo District, Eastern Province, Kenya. Thesis PhD Kenyatta University, Nairobi.

Bless, C., C. Higson-Smith and A. Kagee, 2006. Fundamentals of social research methods: An African Perspective. 1st Edn., Juta and Co, Cape Town.

Buschang, R., G. Chung, G. Delacruz and E. Baker, 2012. Validating measures of algebra teacher subject matter knowledge and pedagogical content knowledge. Educ. Assessment, 17: 1-21.

Carless, D., 2015. Excellence in University Assessment: Leaning from Award-Wining Practice. 1st Edn., Routledge, London.

Eshiwani, G., 2001. Mathematics Education around the world, bridging policy and practice: The role of algebra in the middle and secondary mathematics curriculum. Proceedings of the International Seminar in Park City Mathematics Institute, Jul. 19-23, Utah USA.

Fay, M.P. and M.A. Proschan, 2010. Wilcoxon-Mann-Whitney or t-test? On assumption for hypothesis tests and multiple interpretations of decision rules. Statistics Surveys, 4: 1-39.

Fischer, H., 2011. A History of the Central Limit Theorem: From Classical to Modern Probability Theory, Sources and Studies in the History of Mathematics and Physical Sciences. 1st Edn., Springer, New York.

Gasca, A.M., 2011. Mathematicians and the nation in the second half of the nineteenth century as reflected in the Luigi Cremona correspondence. Sci. Context, 24: 43-72. DOI: 10.1017/S0269889710000268

Gelman, A., 2005. Analysis of variance? Why it is more important than ever. Ann. Statist., 33: 1-53.

Gelman, A., 2008. Analysis of Variance: The New Palgrave Dictionary of Economics. 2nd Edn., Palgrave Macmillan, New York.

Jennison, M. and K. Beswick, 2010. Student attitude, student understanding and mathematics anxiety. Proceedings of the 33rd Annual Conference of the Mathematics Education Research Group of Australasia, (RGA' 10), Fremantle MERGA.

Kleijnen, J.P.C. and R.G. Sargent, 2000. A methodology for the fitting and validation of metamodels in simulation. Eur. J. Oper. Res., 120: 14-29.

Kousha, K. and M. Thelwall, 2008.Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses. J. Am. Society Information Sci. Technol., 59: 2060-2060.

Kutner, M.H., C.J. Nasctsheim, J. Neter and W. Li, 2005.Applied Linear Statistical Models. 5th Edn., McGraw-Hill International, New York.

Makgato, M. and A. Mji, 2006. Factors associated with high school learners' performance: A spotlight on mathematics and physical science. South African J. Educ., 26: 253-266.

Manapure, V., 2011.Effect of further mathematics on students' achievement in mathematics, biology, chemistry and physics. Ind. Streams Res. J., 1: 28-33.

Manoah, S.A., F.C. Indoshi and L.O.A. Othuon, 2011. Influence of attitude on performance of students in mathematics curriculum. Educational Res., 2: 965-981.

Miheso, K.M., 2012. Factors affecting mathematics performance among secondary school students in Nairobi province, Kenya. PhD Thesis, Kenyatta University.

Nelson, R. and P. Dawson, 2014. A contribution to the histort of assessment: How a conversation simulator redeems Socratic method. Assessment Evaluation Higher Education, 39: 195-204.

Newcomer, L., 2009. Universal positive behavior support for the classroom. PBIS Newsletter,

Reed, J.F., F. Lynn and B.D. Meade, 2002. Use of coefficient of variation in assessing variability of quantitative assays. Clin. Diagn Lab Immunol., 9: 1235-1239.

Slattery, J.M. and J.F. Carlson, 2005.Preparing an effective syllabus: Current best practices. College Teaching, 54: 159-164.

Tabachnick, B.G. and L.S. Fidell, 2007.Using Multivariate Statistics. 5th Edn., Pearson Education, Inc. Boston.

Tswani, A.R., 2009. Tracing factors that facilitate achievement in mathematics in traditional disadvantaged secondary schools. PhD Thesis, University of Pretoria, Pretoria.

Umameh, M.A., 2011. A survey of factors responsible for students' poor performance in mathematics in Senior Secondary School Certificate Examination (SSCE) in Idah Local Government Area of Kogi State, Nigeria.

Van Belle, G., 2008. Statistical Rules of Thumb. 2nd Edn., N.J: Wiley, Hoboken.

Wackerly, D., W. Mendenhall and R.L. Scheaffer, 2008.Mathematical Statistics with Applications. 7th Edn., Thomson Higher Education, Belmont.