The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2014

Volume 1, Pages 346-351

ICEMST 2014: International Conference on Education in Mathematics, Science & Technology

CORRELATIONS AMONG ASSESSMENT TECHNIQUES USED IN AN INTRODUCTORY PROGRAMMING COURSE

Cagin KAZIMOGLU
Cyprus International University

Umut TEKGUC
Cyprus International University

Kamil YURTKAN
Cyprus International University

ABSTRACT: Learning computer programming is often perceived to be a difficult task by novice programming students and there have been several studies into the failure rate of students learning to program. This study explores the correlations between introductory programming students' ability to program and their theoretical knowledge in computer programming in order to analyze whether or not their performance in written exams are genuine and accurate. A rigorous statistical analysis was conducted with 400 introductory programming students who were randomly selected without considering whether or not they had a good programming background. The findings of the study used inferential statistics in order to define the correlations between students' practical and theoretical exam results. Additionally, the correlations between students' department and their exam results were investigated in order to define whether or not students' departments have an impact on their success in exams.

Keywords: Introductory programming, computer programming, learning to program, assessment of programming courses

INTRODUCTION

Various studies investigated the reasons why students find computer programming difficult and it was found that the attitudes of students play a key role regarding how difficult they find learning to program (Gomes & Mendes, 2007; Hawi, 2010; Coull & Duncan, 2011). Most introductory programming students perceive computer programming as a technical activity rather than a series of cognitive skills (Bennedsen & Caspersen, 2007). Previous studies sought ways to improve introductory programming courses and investigated the skills of those students who were unsuccessful (Butcher & Muth, 1985; Ford & Venema, 2010). It was found that most students fail to understand the underlying reasons behind the theoretical exams in introductory programming courses where the main concern is not to assess the skill of programming but to measure the ability of abstraction, modelling and debugging (Rajaravivarma, 2005; Dalal *et al.*, 2009).

As a result of these investigations, new instructional design methods were suggested (Ismail *et al.*, 2010; Pears, 2010; Hawi 2012), and the impact of various assessment methods were measured including but not limited to computer aided assessment, multiple choice questions, theoretical/written exams, practical exams, and programming assignments. (Ala-Mutka, 2005; Barros *et al.*, 2003; Chamillard & Braun, 2000; Daly & Waldron, 2004; Lopez *et al.*, 2008; Kuechler & Simkin, 2003).

Despite the encouraging work done in this area, to this day it is still arguable whether or not students comprehend the motivation behind the concept of abstraction, modelling and debugging and as a result of this, they do not always see the point of theoretical exams in introductory programming courses. Consequently, there is the common assumption that students who are studying a Computer Science degree tend to do better in computer programming than students who are not studying a Computer Science degree.

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Cagin KAZIMOGLU- e-mail: ckazimoglu@ciu.edu.tr

This paper is dedicated to measure these assumptions and in order to do this, a rigorous evaluation was undertaken to determine whether or not there is a significant relationship a) between students' theoretical and practical exam results; b) among students' departments and their theoretical and practical exams. It was aimed to measure whether or not there is a significant correlation between students' results in their theoretical exams with the results they obtained in their practical exams. Both of these aspects are analyzed separately and in combination, to ensure the accuracy of the results and the benefits that can be derived from this.

EXPERIMENTAL DESIGN

A rigorous study was undertaken to compare theoretical exam results of students with their practical exam results in order to measure whether or not these results are correlated. These results were gathered from the Faculty of Engineering students who were in their first year. Due to the engineering formation and school policy, all engineering students are obligated to take the introductory to programming course at the Cyprus International University (CIU). There are nine different departments under the Faculty of Engineering at CIU which are Computer Engineering, Information System Engineering, Electric and Electronic Engineering, Industrial Engineering, Energy Engineering, Civil Engineering, Environmental Engineering, Management Information Systems and finally Computer Programming. Regardless of their departments, students received computer algorithms for the first couple of weeks during their introductory programming course where they were thought to draw flowcharts and wrote pseudo codes in order to develop their ability to think computationally. Having covered fundamentals of computer algorithms, students were introduced the C programming language and asked to attend two sets of theoretical and practical exams along the course. The first theoretical exam results are matched with the first practical exam results and the second theoretical exam results with the second practical exam results. The first set of exams were related to the structured computer programming concepts in C covering the most basic computer programming constructs: variables, sequence, decision making and loops. Consequently, the second sets of exams were linked to the modular computer programming specifically focusing on pre-defined and user-defined functions as well as arrays.

A total of 400 comparable valid results from theoretical and practical exams were gathered and entered into IBM software package used for statistical analysis (SPSS). The results of the exams were gathered randomly without considering whether or not students had prior knowledge or background in programming. The identities of students were kept confidential and only the departments of the students were used in the statistical analysis as nominal data. The theoretical and practical exam results were used as raw data to investigate the validity of the following research questions and their hypotheses:

#	Research Question	Null Hypothesis (H_o1)	Alternative Hypothesis (H_a1)
1	Is there a significant relationship between student's theoretical and practical exam results?	There is no significant relationship between students' theoretical and practical exam results.	Students' theoretical exam results are significantly related with their practical exam results.
2	Is there a significant relationship between students' theoretical exam results and their departments?	There is no significant relationship between students' theoretical exam results and their department.	There is a significant relationship between students' theoretical exam results and their department.
3	Is there a significant relationship between students' practical exam results and their departments?	There is no significant relationship between students' practical exam results and their department.	There is a significant relationship between students' practical exam results and their department.

In order to examine the results accurately in the context of the above research questions, it was crucial to identify the correct method for an inferential statistical analysis. As the experimental structure is based on investigating the relationship of theoretical and practical exam results in a sample group, a statistical hypothesis was needed to evaluate the correlations on the subject. Therefore, a procedure for carrying out either a Pearson's product-moment correlation coefficient or a Spearman's rank-order correlation was performed (henceforth referred to as Pearson's r and Spearman's correlation). A Pearson's r was to be selected should the data captured fit a normal distribution and similarly, Spearman's correlation was to be available if the data captured did not fit a normal distribution.

When working with regional data (i.e. exam results) that comes from a normally distributed population, a Pearson's r is used to identify the strength and direction of correlations between the variables. Similar to this, a Spearman's correlation is selected when data comes from a non-normal distribution. Both methods can be used to measure how strong a correlation is between two variables. However, in Spearman's correlation it is simply not possible to calculate the percentage of variance as a coefficient of determination whereas this can be calculated in Pearson's r. Therefore, having entered the raw data into SPSS, the procedure for checking the normal distribution was undertaken in order to decide which statistical method to use for the analysis of data.

STATISTICAL ANALYSIS

Three different methods were used in order to identify the distribution of data: *Histogram, Quantile – Quantile Plots* and a *Skewness and Kurtosis normality check*.

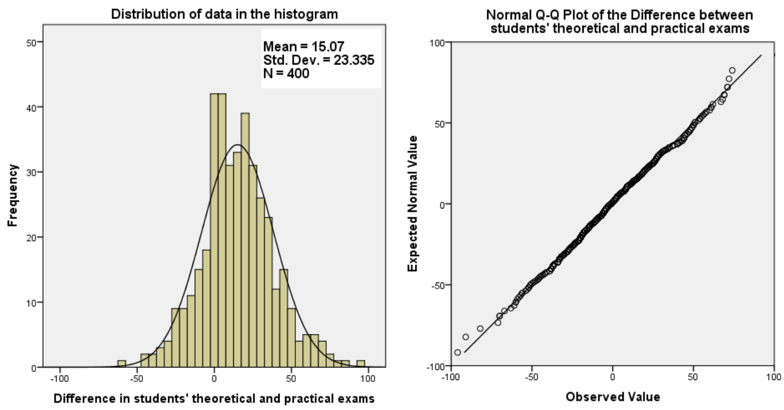

Figure 1. Normal Distribution Of Data In The Histogram And In The Normal Quantile – Quantile (Q-Q) Plot.

As shown in Figure 1, a histogram and a Quantile – Quantile (Q-Q) plot was used as the predominantly methods for observing how close the distribution of data is to a normal distribution. The histogram shows that there are skewness issues in the distribution of data as it is skewed to the right which causes an asymmetry in the distribution. Additionally, the histogram has kurtosis issues as the peak kurtosis point is over than what would be expected in a normal curve. Hence, there are clusters in the distribution of data which suggests that the data came from a non-normally distribution. On the other hand, the Q-Q plot shows that the differences in the exam results hug the linear line and no multiple clusters are visible which proves that the data obtained does not concentrate on specific points. In other words, the Q-Q plot provides strong reasons to believe that the data came from a normally distributed population.

Despite the strong evidence obtained from Q-Q plots, it was necessary to define the degree of Skewness and Kurtosis issues in terms of statistics, as the histogram did not support the results in Q-Q plots.

Table 1. Skewness And Kurtosis Normality Check On The Difference Between Practical And Theoretical Exams.

	N	Mean		Stdandard Deviation	Skewness		Kurtosis	
		Statistics	Std. Error		Statistics	Std. Error	Statistics	Std. Error
Difference between Pract. and Th. midterm	400	15.07	1.167	23.335	**.186**	.122	**.563**	.243
Difference between Pract. and Th. final	400	20.80	1.167	23.347	**.118**	.122	**.298**	.243
Valid N (listwise)	400							

As shown in Table 1, the Skewness and Kurtosis values are very close to 0. Moreover, the standard error calculation for Skewness (0.122*3=0.366>0.186) and Kurtosis (0.243*3=0.729>0.563) satisfy that the data came from a normally distributed population.

The above statistical procedures (i.e. histogram, Q-Q plots and Skewness and Kurtosis normality check) were discussed for the first research question only. Because the data set obtained for each research question has gone through the same normality tests, only the procedure for the first question is described here as the rest of the research questions were analyzed in the same way.

A Pearson's r was computed to assess the relationships among the exams (i.e. midterm and final) and their relationship to departments. As shown in Table 2, the correlations between the theoretical and practical exams are in positive direction, moderately strong and significant. There is a positive, modestly strong and significant association in between theoretical and practical midterm exam ($r=0.621$, $n=400$, $p=0.01$) and in between theoretical and practical final exam ($r=0.651$, $n=400$, $p=0.01$). This means that the Pearson's coefficient provides strong evidence that the associations are moderately strong ($r^2=0.42$, 42%) between theoretical and practical exams.

Table 2. Pearson Product-Moment Correlation Coefficient showing relationships between Practical and Theoretical Exams.

**. Correlation is significant at the 0.01 level (2-tailed).		Theoritical Exam Midterm	Practical Exam Midterm
Theoritical Exam Midterm	Pearson Correlation	1	**.621****
	Sig. (2-tailed)		.000
	N	400	400
Praticial Exam Midterm	Pearson Correlation	**.621****	1
	Sig. (2-tailed)	.000	
	N	400	400

		Theoritical Exam Final	Practical Exam Final
Theoritical Exam Final	Pearson Correlation	1	**.651****
	Sig. (2-tailed)		.000
	N	400	400
Praticial Exam Final	Pearson Correlation	**.651****	1
	Sig. (2-tailed)	.000	
	N	400	400

Based on the analysis of data, two different conclusions can be drawn from the Pearson's r. Firstly, those students who did well in their midterm exams also did reasonably well (or even better) in their final exams as the difference between the coefficient number is very small (difference in $r = 0.03$, $p=0.01$). Secondly, there is no significant or strong difference between students' theoretical and practical exam results. There is a modestly strong correlation between the theoretical and practical exams in both cases which provides strong reasons to believe that those students who did well in their theoretical exams also did well in their practical exams. Hence, there is strong and significant evidence to support the alternative hypothesis for the first research question. In other words, the results provide strong reasons to believe that students' theoretical exam results are significantly related with their practical exam results.

One-way Multivariate analysis of variance (MANOVA) is used to determine whether there are any differences between the independent groups (i.e. students' department) on more than one continuous dependent variable (i.e. theoretical and practical exams, both midterm and final). At this point, it is crucial to note that the one-way MANOVA is an omnibus test and thus do not recognize which specific groups were significantly different from each other. In other words, it can only show that at least two groups were significantly different from each other. Because of this reason, a post-hoc test (i.e. Tukey's range test) was used but the results of this are not demonstrated here due to the lack of space. Additionally, MANOVA has a series of prior assumptions (e.g. no multivariate outliers, adequate sample size) and the validity of these were in order to ensure that the data obtained would not violate the results.

The MANOVA test included a single independent variable (i.e. students' department) and four dependent variables (students' theoretical midterm exams results, students' practical midterm exams results, student's theoretical final exam results and finally students' practical final exam results). The outcomes regarding the multivariate tests are given below:

Table 3. The Multivariate Tests[a] table showing the second Effect, labelled "Department", and the Wilks' Lambda row

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.166	38.901[b]	2.000	390.000	.000
	Wilks' Lambda	.834	38.901[b]	2.000	390.000	.000
	Hotelling's Trace	.199	38.901[b]	2.000	390.000	.000
	Roy's Largest Root	.199	38.901[b]	2.000	390.000	.000
Department	Pillai's Trace	.034	.842	16.000	782.000	.638
	Wilks' Lambda	**.966**	**.840[b]**	**16.000**	**780.000**	**.640**
	Hotelling's Trace	.034	.838	16.000	778.000	.642
	Roy's Largest Root	.019	.919[c]	8.000	391.000	.500

a. Design: Intercept + Department b. Exact statistic c. The statistic is an upper bound on F that yields a lower bound on the significance level.

As shown in Table 3, there was no statistically significant difference in student's performance with regard to their departments, $F (16, 780) = 0.84$, $p > .0005$; Wilk's $\Lambda = 0.966$. To determine whether the one-way MANOVA was statistically significant it was required to check the "Sig." value on the multivariate tests table, which in this case is 0.640. As the significant value is more than 0.005, it is possible to conclude that the alternative hypotheses for the second and third research questions can be rejected. In other words, there is no strong or significant evidence to support that students' performance in their theoretical and practical exams was related to which department they came from. Hence, the results of the one-way MANOVA provides strong reasons to believe that students' practical and theoretical exam results were not associated on which department they studied in.

CONCLUSION

As discussed earlier in the paper, majority of students tend to believe that practical exams have a higher impact of assessment than theoretical exams. Additionally, there is a common assumption that students who are studding a computer science degree perform better in computer programming than other engineering students. This study investigated how closely practical exams and theoretical exams were related to each other as well as whether or not the department of students has an impact on their performances. The statistical results gathered from this study provided strong reasons to believe that there is a significant relationship between student's theoretical and practical exam results. The results also show strong reasons to believe that there is no significant relationship between the students' department and their practical and/or theoretical exam results.

The information gathered during this study is only based on the overall results gathered from students' practical and theoretical exams. This study can further be expanded by analyzing the relationship of practical and theoretical exam results in terms of different programming constructs (such as loops, decision making, and functions). Therefore, it could be possible to detect which assessment methods work better on measuring which programming constructs.

REFERENCES

Ala-Mutka, K. M. (2005). A survey of automated assessment approaches for programming assignments. Computer Science Education, 15(2), 83-102.

Barros, J. P., Estevens, L., Dias, R., Pais, R., & Soeiro, E. (2003, June). Using lab exams to ensure programming practice in an introductory programming course. In ACM SIGCSE Bulletin (Vol. 35, No. 3, pp. 16-20). ACM.

Bennedsen, J., & Caspersen, M. E. (2007). Failure rates in introductory programming. ACM SIGCSE Bulletin, 39(2), 32-36.

Butcher, D. F., & Muth, W. A. (1985). Predicting performance in an introductory computer science course. Communications of the ACM, 28(3), 263-268.

Chamillard, A. T., & Braun, K. A. (2000). Evaluating programming ability in an introductory computer science course. ACM SIGCSE Bulletin, 32(1), 212-216.

Coull, N. J., & Duncan, I. M. (2011). Emergent requirements for supporting introductory programming. Innovation in Teaching and Learning in Information and Computer Sciences, 10(1), 78-85.

Dalal, N., Dalal, P., Kak, S., & Antonenko, P. (2009). Rapid digital game creation for broadening participation in computing and fostering crucial thinking skills. International Journal of Social and Humanistic Computing, 1(2), 123-137.

Daly, C., & Waldron, J. (2004, March). Assessing the assessment of programming ability. In ACM SIGCSE Bulletin (Vol. 36, No. 1, pp. 210-213). ACM.

Ford, M., & Venema, S. (2010). Assessing the Success of an Introductory Programming Course. Journal of Information Technology Education, 9.

Gomes, A., & Mendes, A. J. (2007). Learning to program-difficulties and solutions. In International Conference on Engineering Education–ICEE (Vol. 2007).

Hawi, N. (2010). Causal attributions of success and failure made by undergraduate students in an introductory level computer programming course. Computers & Education, 54(4), 1127-1136.

Hawi, N. S. (2012). A CAD (Classroom Assessment Design) of a Computer Programming Course. Online Submission.

Ismail, M. N., Ngah, N. A., & Umar, I. N. (2010). INSTRUCTIONAL STRATEGY IN THE TEACHING OF COMPUTER PROGRAMMING: A NEED ASSESSMENT ANALYSES. Turkish Online Journal of Educational Technology, 9(2).

Kuechler, W. L., & Simkin, M. G. (2003). How well do multiple choice tests evaluate student understanding in computer programming classes?. Journal of Information Systems Education, 14(4), 389-400.

Lopez, M., Whalley, J., Robbins, P., & Lister, R. (2008, September). Relationships between reading, tracing and writing skills in introductory programming. In Proceedings of the Fourth international Workshop on Computing Education Research (pp. 101-112). ACM.

Pears, A. N. (2010, October). Enhancing student engagement in an introductory programming course. In 40th Frontiers in Education Conference, ser. Proceedings of the Frontiers in Education Conference (No. 40).

Pears, A., Seidman, S., Malmi, L., Mannila, L., Adams, E., Bennedsen, J., ... & Paterson, J. (2007, December). A survey of literature on the teaching of introductory programming. In ACM SIGCSE Bulletin (Vol. 39, No. 4, pp. 204-223). ACM.

Rajaravivarma, R. (2005). A games-based approach for teaching the introductory programming course. ACM SIGCSE Bulletin, 37(4), 98-102.