

A quantitative assessment of student performance and examination format

Christopher B. Davison
Ball State University

Gandzhina Dustova
Ball State University

ABSTRACT

This research study describes the correlations between student performance and examination format in a higher education teaching and research institution. The researchers employed a quantitative, correlational methodology utilizing linear regression analysis. The data was obtained from undergraduate student test scores over a three-year time span. The purpose of this study was to investigate the predictive relationships between standardized examinations and practical examinations. The data consists of 247 undergraduate students' test scores spanning three academic years. Computer Technology students were assigned to take a standard midterm exam as well as a practical exam. The result of the analysis demonstrates that standardized examination scores are not predictors of practical examination scores and may well be testing different skill sets.

Keywords: Standard exam, practical exam, test score, assessment, predictive modeling.

INTRODUCTION

This research study determined if any correlations exist between student performance and examination format in a large, Midwestern research/teaching institution. The study data was derived from student examination performance scores. The data was collected from two technology-related courses over a three-year timeframe.

In this quantitative, correlational study using regression analysis, a predictive model was created for each course. The research question proposed for this study is: are the standard examination scores a good predictor of the practical (i.e., hands-on) examination scores.

Department of Technology faculty members noticed that there is a significant student performance differential between the standard examination and practical examination formats. Students who do well in the standard examination do not necessarily perform well in the practical examination. Resultant from this observation, the correlation and predictive modeling between the examination types were studied.

Purpose of the study

The purpose of this is to examine the relationship between the standard examinations (typical True/False and multiple-choice questions) and practical examinations (hands on system administration tasks) for undergraduate students in a Midwestern computer technology program. The program is a part of the Department of Technology at a large research and teaching university.

Research Question

Are scores from standard examinations good predictors of performance on practical (hands on) examinations? To test this research question, data from two technology related courses were analyzed. The data was obtained from three years of test scores from a 200-level Systems Administration course and a 300-level Infrastructure Services course.

Hypothesis

Null Hypothesis (H_{10}): The midterm standard examination score does not significantly predict the midterm practical examination score for undergraduate students in a Midwestern computer technology program.

Alternative Hypothesis (H_{1A}): The midterm standard examination score does significantly predict the midterm practical examination score for undergraduate students in a Midwestern computer technology program.

Both the Null and Alternative hypothesis were tested for two courses. The first course was a 200-level computer technology course focusing on systems administration. The second course was a 300-level computer technology course focusing on infrastructure services.

Variables

The independent variable selected for this study is midterm standard examination. This variable was selected as a predictor for the dependent variable. The standard examination

consists of a mix of 25 true or false and multiple-choice questions focused on MS Windows Server systems administration. Each question is worth 2 points for a total of 50 points. The majority of the test questions are derived from the textbook publisher's test-bank that is derived from the Microsoft 70-410 certification examination.

The practical exam is a series of 8 MS Windows Server systems administration tasks. Each task is weighted between 10-20 points per successful outcome for an overall possible score of 100 points. The practical examination tasks are derived from the textbook material and closely related to the standard exam questions.

The context of all of the questions and systems administration tasks was the Microsoft certification examinations. Specifically, the Exam 70-410 Microsoft Official Academic Course for the System Admin Fundamentals (TCMP 211) Course and the Exam 70-412 Microsoft Official Academic Course for the Infrastructure Services (TCMP 311) Course. These courses prepare students for additional certification exams in their field of study for career enhancement. Additionally, these two courses are required for the Bachelor's Degree in Computer Technology.

Environment and Control

Both the practical exam and the standard exam take place in a classroom. The time limit for both examinations is 75 minutes. All students have finished both examinations within the time allotted. No additional time was required or requested by the students in any testing phase over the course of the data collection period.

The standardized exam is administered through the Blackboard system. Students open a web browser, login to the course room, and then take the examination. The Blackboard system scores the examination when the student submits it and immediately returns the score.

The instructor administers the practical examination. All systems administration tasks are projected on a screen along with their concomitant point value (10-20 points per task). The students select the tasks and the order in which the tasks are attempted. The students provide screen shots of the tasks attempted or completed. All of the tasks are performed on a pre-configured Windows Server 2012 virtual machine. Each student is provided a workstation with the working virtual machine installed on it.

Both of the exams were administered in the same week at the same time of day. Both courses meet twice a week at the same time for 75 minutes. The standardized exam was administered on the first course meeting during Midterm week. The practical exam was administered two days later.

Timeframe of Data Collection

The data collection period was three years. The data was analyzed for correlations using SPSS software package.

BACKGROUND LITERATURE

Over the past two decades, there has been an upsurge of interest in how achievement goals influence self-regulated learning and academic performance (Covington, 2000). There are number of existing studies pertaining to academic performance and factors that contribute to academic performance. Teacher engagement and student motivation are large areas of research

in this domain (Zimmerman, Schmidt, Becker, Peterson, Nyland & Surdick, 2014). Additionally, there exists pedagogical research comparing standard examinations to practical examinations (Davison, 2015). However, there appears to be a gap in the research literature with regard to using standard examination scores as a predictor of practical examination scores. In this research article, this gap in the research literature is addressed by creating two predictive models (one per course) using standard examination scores as the independent variable and practical examination scores as the dependent variable.

Academic achievement (i.e., GPA or grades) is one tool to measure students' academic performance. Based on the Center for Research and Development Academic Achievement (CRIRES) (2005) report, academic achievement is a construct to measure students' achievement, knowledge and skills. This measurement is holistically based on the students' age, the students' previous experience, and the students' capacity related to social and education skills. To measure academic achievement, educators use different types of assessment. Assessment is a continuous process that brings some valuable information about the learning process (Linn and Gronlund, 1995). Hargis (2003) commented that the grading process is supposed to be motivating and provide goals. On the other hand, grades can provide incentives to the students to cheat. Grading has the additional benefit of provide records (data sets) of students' academic achievements. (Haladyna, 1999).

Factors such as confidence (Schunk, 1991), and motivation (Covington, 2000; Kohn, 1993; Stiggins, 2001; Tuckman, 1998) influence students' ability to score well on exams. According to Siang & Santoso (2016), educators have a number of tools at their disposal to assist students. With regard to these tools, "perhaps the most entrenched strategy is that of tests and grades, which operate in a punishment–reward fashion" (Myers & Myers, 2007, p. 227). However, the efficacy of exams, from the classroom to college admissions, is debated and controversial (Linn, 2001).

In the usual lecture/lab form of classroom instruction, midterms and final examinations are common. However, a large number of researchers criticize these examinations formats as not conducive to retaining information and student inclination to *cram* (Donovan & Radosovich, 1999; Willingham, 2002). A large body of research literature encourages alternative testing strategies to better support student achievement and information retention (Bahji, Lefdaoui, & Alami, 2013; Chen, & Liao, 2013).

With regard to the alternative testing strategies, the purpose of this study was to perform a qualitative assessment of student performance versus examination format. Two assessment methods of academic achievement among undergraduate students enrolled in two computer technology courses were applied: a standard midterm examination structure and a practical (hands-on) examination. The hypothesis guiding this research is that one examination format is correlated to the other and could serve as a predictor.

There are a number of studies that examine correlations in examination formats and quizzes. Haberyan (2003) studied undergraduate students and found no statistical correlation between weekly quizzes and examinations. Graham (1999) found that psychology undergraduates performed better on examinations when subject to random quizzes throughout the semester. Furthermore, the lower GPA achieving students tended to benefit the most from the random quizzes.

In the Ruscio (2001) research, random quizzes were administered in order to test whether the students were performing the assigned reading. The result from this research indicates that students achieving high quiz scores (because of performing the required reading) tended to do

better on the other types of course assessments. Relatedly, Tuckman (1996, 1998) promotes a multi-examination strategy to increase overall test scores and promote more studying.

According to Myers and Myers (2006) the effects of different examination formats on student GPA scores are not precisely known. They do suggest that GPA score is higher when the frequency of examinations are higher (bi-weekly as opposed to one midterm examination). The studies that do focus on this area tend to be more short-term and do not track student achievement over time. More longitudinal work in this research domain is necessary.

METHODOLOGY AND DESIGN

The research design selected for this study is a quantitative methodology utilizing a correlational study design. Creswell (2005) encourages this design in order to produce predictive models. In explaining correlation research, Shirish (2013) states, “this design is appropriate as correlational research attempts to determine the extent of a relationship between two or more variables using statistical data” (p. 71). It is important to note that a correlation between variables is not necessarily causality.

The purpose of the study is to examine relationships (if any) between standardized test scores and practical exam scores. As one of the outcomes from this study is a predictive model, the research design utilized linear regression analysis. This design type also allows for hypothesis testing. The methodology selection was driven by the research question.

Data Collection

The data was obtained from 247 undergraduate exam scores in the department. The data was stored in the Blackboard system and retrieved for the purposes of this research. The data was analyzed using the SPSS statistical package. Resultant predictive models were derived from the SPSS analysis.

RESULTS AND DISCUSSION

Data from two TCMP System Administration courses (TCMP211, TCMP311) was analyzed. The data sets consist of several years' worth of two Midterm examination types: Practical Assessments and Standardized Examination (e.g., True/False questions, Multiple Choice questions). The data from those examinations was analyzed in terms of correlations and score prediction. Findings presented are aggregate findings from course scores over a three-year timeframe.

The findings suggest that the average score for the 200-level standardized test is 73% (2.0 GPA). The practical exam average in that course is 76% (2.0 GPA) (see Table 1 in the Appendix). The practical exam does have an interestingly high standard deviation at 20, while the standard exam only has a standard deviation of 6.

In the 300-level course data set, the average score is 67% (1.3 GPA) for the standard exam. The practical exam has a much higher average score at 84% (3.0 GPA). For the standard deviations, the 300-level course data indicates a 24 for the practical exam and 8 for the standard.

Next, the overall score (final grade and GPA) for students was analyzed. The range of course GPAs for the TCMP 211 course is .13 to 3.975. The range of course GPAs for the TCMP 311 course is .28 to 3.88.

As presented above, the standard deviation for the practical assessment (20) is much higher than the standard test (6) as is the Variance (379 vs. 33) in TCMP 211. Likewise, in TCMP 311 the standard deviation is 8 in the standard exam and 24 in the practical exam and the Variance is 64 and 571 respectively. This suggests a high degree of variation in the two sets of test scores. This could be partially attributed to a higher spread in the MIN and MAX scores between the two exams. However, much of this is caused by a significant amount of low scores and high scores in the practical examination. This would indicate that students taking the practical are either extremely proficient with regard to the course material or they are not.

The predictive model used the standard midterm examination as a predictor of the midterm practical examination score. In both TCMP 211 and TCMP 311 the models experienced a very high standard error of the estimate (see Table 2 in the Appendix). Relatedly, the R^2 for both courses was very close to 0. This indicates that student results on the standardized midterm exam is not a predictor of their ability to perform on the practical midterm. The practical exam and the standard exam are measuring separate skill sets.

For scientific purposes, the regression equations (e.g., predictive models) are presented for both courses. As previously stated, each model suffers from low R^2 values so the goodness-of-fit of the values is poor. Relatedly, the TCMP211 regression equation is not statistically significant (.073) while the TCMP311 regression equation is significant (.001) (see Table 3 in the Appendix).

Predictive Model for TCMP211:

$$y = 57.572 + .516(x)$$

Where

$y =$ TCMP211 Practical Exam score ($100 \geq y \geq 0$)

and

$x =$ TCMP 211 Standard Exam score ($50 \geq x \geq 0$)

Predictive Model of TCMP311:

$$y = 51.55 + .969(x)$$

Where

$y =$ TCMP 311 Practical Exam score ($100 \geq y \geq 0$)

and

$x =$ TCMP 311 Standard Exam score ($50 \geq x \geq 0$)

Impact of Results on Hypotheses

For the TCMP 211 course, the Null hypothesis could not be rejected. For the TCMP 311 course, the Null hypothesis can be rejected, resulting in a statistically significant predictive model presented earlier. However, in both cases, the R^2 was close to 0 (see Appendix, Table 2). This means that the resultant model (while statistically significant for the TCMP 311 course) is not a good fit as the model suffers from high unexplained variance.

CONCLUSION

This research study explored the relationships of student scores from practical and standard type of examinations. The methodology employed was a quantitative, correlational

approach utilizing linear regression analysis to describe any predictive relationship between the examination types. The results indicate that both predictive models (for the 200-level course and the 300-level course) suffer from a high degree of unexplained variance. As such, the predictive value of the standardized examination score in relation to the practical examination score is low. While the resultant model was statistically significant for the 300-level course, the usefulness of this model is limited due to the very low R^2 value.

Based on the results of the data analysis, it appears that within the sample set the standardized examinations are testing different skill sets than the practical examinations. The students' ability to answer True/False and multiple-choice questions regarding the subject material is not a good predictor of the ability to apply the subject material in a hands-on, practical fashion. This observation is limited to two courses that are required computer technology specific courses.

This research is exploratory in nature and was specifically limited to the undergraduate students in a large, public, Midwestern computer technology program. The results provided a deeper insight into examination types and could assist educators in selecting a type of examination to administer to their students.

REFERENCES

- Bahji, S.E., Y. Lefdaoui, and J. El Alami. (2013). Enhancing Motivation and Engagement: A Top Down Approach for the Design of a Learning Experience According to the S2P-LM. *International Journal of Emerging Technologies in Learning*, 8(6).
- Center for Research and Development Academic Achievement (CRIRES) (2005). Data taken from International Observatory on Academic Achievement. Retrieved from http://www.crires-oirs.ulaval.ca/sgc/lang/en_CA/pid/5493
- Chen, M.H, Liao, J.L. (2013). Correlations among Learning Motivation, Life Stress, Learning Satisfaction, and Self-Efficacy for Ph.D Students. *The Journal of International Management Studies*, 8(1), 157 – 162.
- Creswell, J. (2005). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (2nd Ed.). Upper Saddle River, New Jersey: Pearson.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, 51, 171–200.
- Davison, C.B. (2015). Assessing IT Student Performance Using Virtual Machines. *Tech Directions*, 74(7), 23-25.
- Hargis, C.H. (2003). *Grades and Grading Practices. Obstacles to Improving Education 114 and to Helping At-Risk Students* (2nd ed.) Springfield, IL: Thomas.
- Haladyna, T. M. (1999). *A Complete Guide to Student Grading. Needham Heights. MA: Allyn & Bacon.*

- Linn, R. L. (2001). A century of standardized testing: Controversies and pendulum swings. *Educational Assessment*, 7(1), 29-38.
- Linn, R.L. & Gronlund, N.E. (1995). *Measurement and Evaluation in Teaching*, (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Myers, C.B. & Myers, S.M. (2006). Assessing Assessment: The Effects of Two Exam Formats on Course Achievement and Evaluation. *Innovative Higher Education*, 31(4), 227-236.
- Siang, J. J., & Santoso, H. B. (2016). Learning Motivation And Study Engagement: Do They Correlate With Gpa? An Evidence From Indonesian University. *Researchers World : Journal of Arts, Science and Commerce RW-JASC*, 7(1(1)), 111-118.
doi:10.18843/rwjasc/v7i1(1)/12
- Schunk, D.H. (1991). Self-efficacy and Academic Motivation. *Educational Psychologist*, 26, 207-231.
- Shirish, T.S. (2013). *Research Methodology in Education*. USA: Lulu.
- Tuckman, B. W. (1996). The relative effectiveness of incentive motivation and prescribed learning strategies in improving college students' course performance. *The Journal of Experimental Education*, 64, 197-210.
- Tuckman, B. W. (1998). Using tests as an incentive to motivate procrastinators to study. *The Journal of Experimental Education*, 66, 141-147.
- Zimmerman, T., Schmidt, .L, Becker, J., Peterson, J., Nyland, R., & Surdick, R. (2014). Narrowing the Gap between Students and Instructors: A Study of Expectations. *Transformative Dialogues: Teaching and Learning Journal*, 7(1), 1-18.

APPENDIX

Table 1.

TCMP 211 Descriptive Statistics

	Mean	Std. Deviation	N
Midterm_Practicum [Total Pts: 100] 1551307	76.47	19.462	139
MidTerm [Total Pts: 50] 1551316	36.65	5.761	139

TCMP 311 Descriptive Statistics

	Mean	Std. Deviation	N
Midterm Practicum [Total Pts: 100] 891197	84.13	23.897	108
MidTerm [Total Pts: 50] 891196	33.74	7.996	108

Table 2.

TCMP 211 Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.153 ^a	.023	.016	19.304

a. Predictors: (Constant), MidTerm [Total Pts: 50] |1551316

b. Dependent Variable: Midterm_Practicum [Total Pts: 100] |1551307

TCMP 311 Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.324 ^a	.105	.097	22.713

a. Predictors: (Constant), MidTerm [Total Pts: 50] |891196

b. Dependent Variable: Midterm Practicum [Total Pts: 100] |891197

Table 3.

TCMP 211 Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	57.572	10.581		5.441	.000
MidTerm [Total Pts: 50] 1551316	.516	.285	.153	1.808	.073

a. Dependent Variable: Midterm Practicum [Total Pts: 100] |1551307

TCMP 311 Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	51.440	9.520		5.404	.000
MidTerm [Total Pts: 50] 891196	.969	.275	.324	3.528	.001

a. Dependent Variable: Midterm Practicum [Total Pts: 100] |891197

