Gaming and Surgeon Performance

Jessica L Barron

Duquesne University

Data Analysis Project

GREV 610

Gibbs Y. Kanyongo, PhD

April 30, 2011

Introduction

Video games have become the topic of discussion in education and have been the focus of many studies involving teaching, learning and the application of specific skills. The research for this study will examine how students learning to become laparoscopic surgeons perform in simple tasks, such as laparoscopic navigation and suturing skills. The significance of the study will bring awareness to different instruction techniques that will enhance hand eye coordination and fine motor skills. Inspiration for this study came from a similar paper titled “The impact of video games in surgical training” (Rosser, J. C., Lynch, P. J. Haskamp, L., Gentile, D. A., & Yalif, A., 2007).

A one-way repeated-measures analysis of variance will be conducted to determine whether gameplay in three different video games will affect the error rate in the Barron Suturing Test. The Barron Suturing Test assesses the learner’s skill in following a set course with the laparoscopic scope that ends with a specific amount of sutures they must complete. The test is timed and a score is then given to the learner. The test will be given at the start of their schooling, and then repeated three times afterwards every six months. The idea is that the gameplay will give the learners ability to learn and memorize a course, repeat specific commands and complete them within a time limit.

Three games with this criterion will be Wii Sports Resort, Mario Kart, and Super Mario Brothers. All three games are played on the Nintendo Wii. Each game forces the player to learn specific commands and complete a course or level within a set time limit. For the first six months, the subjects will complete at least three hours of gameplay a week. For the next six months, the hours will be increased to six. The last six months will be set at nine hours.

The population will be 30 resident surgeons entering their schooling in a local academic medical center and surgical training program. The surgeons are at the same academic level and each one is completing their schooling to become a laparoscopic surgeon. These surgeons have been picked because of their similar academic and skill abilities. The scores they achieve have no dependency upon the next score; they will not be competing with one another. Their scores will be based on errors made within the test.

The hypothesis is that these games will increase performance on the Barron Suturing Test over an 18 month period.

The Data

Data will be collected four times using the Barron’s Suturing Test and its scoring system. Score will be based upon time and the amount of errors that occur within the test. One point will be deducted for every error and also for every minute over the time limit. The maximum score will be set at 100 points.

The first test will be upon entry of the program. Then the test will be repeated three more times; once at six months, again at twelve months and finally at eighteen months.

When using a one way repeated measures analysis of variance, a sample size of 30 is acceptable if the population distribution is considered normal. Since there are no significant variances amongst the population, a sample of 30 surgeons was used.

In order to obtain the numbers, two methods were used. First, the file from the lesson 29 data sheet was used for formatting purposes and was altered to fit the needs of the study. Then the data generator program was used to create a study with four variables. The random numbers were generated with specific means with the minimum score set to 1 and the maximum to 100. When the data was generated, the numbers were imported into the altered data sheet.

Results

The following output was used to analyze the results of the test and shown in figures 1.1, 1.2 and 1.3. The means are shown in figure 1.1. The standard Univariate ANOVA, shown in figure 1.3 under Sphericity Assumed, indicates a significant time effect, F(3,87) = 1754, p<.01. The alternate univariate test (Greenhouse-Gessier and Huynh-Feldt) yield the same F value with the correct degrees of freedom. The effect for these tests is P<.01 as well. The Wilks Lambda = .004, F(3,27) = 2255, p <.01.

|  |  |  |  |
| --- | --- | --- | --- |
| **Figure 1.1**  **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| Score at month 1 | 59.70 | 1.643 | 30 |
| Score at month 6 | 69.60 | 1.248 | 30 |
| Score at month 12 | 79.40 | 2.283 | 30 |
| Score at month 18 | 89.83 | 1.147 | 30 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Figure 1.2**  **Multivariate Testsb** | | | | | | | |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| errorscore | Pillai's Trace | .996 | 2255.112a | 3.000 | 27.000 | .000 | .996 |
| Wilks' Lambda | .004 | 2255.112a | 3.000 | 27.000 | .000 | .996 |
| Hotelling's Trace | 250.568 | 2255.112a | 3.000 | 27.000 | .000 | .996 |
| Roy's Largest Root | 250.568 | 2255.112a | 3.000 | 27.000 | .000 | .996 |
| a. Exact statistic  b. Design: Intercept  Within Subjects Design: errorscore | | | | | | | |

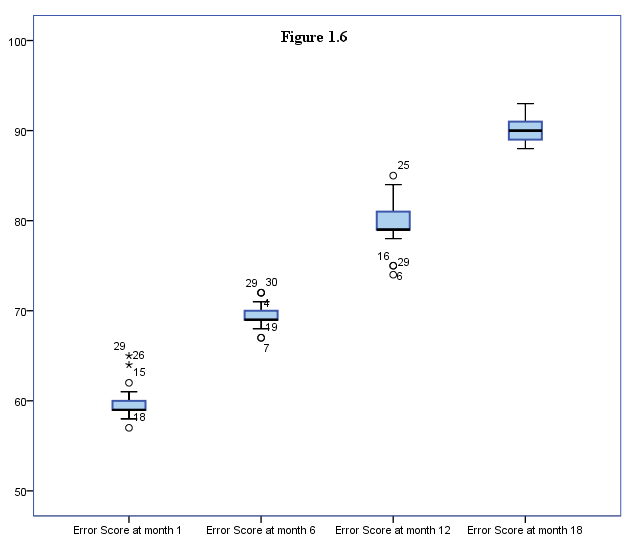
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Figure 1.3**  **Tests of Within-Subjects Effects** | | | | | | | |
| Measure:MEASURE\_1 | | | | | | | |
| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
| errorscore | Sphericity Assumed | 15063.000 | 3 | 5021.000 | 1754.325 | .000 | .984 |
| Greenhouse-Geisser | 15063.000 | 2.075 | 7260.071 | 1754.325 | .000 | .984 |
| Huynh-Feldt | 15063.000 | 2.237 | 6732.291 | 1754.325 | .000 | .984 |
| Lower-bound | 15063.000 | 1.000 | 15063.000 | 1754.325 | .000 | .984 |
| Error(errorscore) | Sphericity Assumed | 249.000 | 87 | 2.862 |  |  |  |
| Greenhouse-Geisser | 249.000 | 60.168 | 4.138 |  |  |  |
| Huynh-Feldt | 249.000 | 64.885 | 3.838 |  |  |  |
| Lower-bound | 249.000 | 29.000 | 8.586 |  |  |  |

Next, the pairwise comparisons will be analyzed in Figure 1.4. The test scores will be compared to one another in order to find significance. The p value for all of the comparisons are significant p = <.01 Therefore, there is a notable difference between all of the scores obtained.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Figure 1.4**  **Pairwise Comparisons** | | | | | | |
| Measure:MEASURE\_1 | | | | | | |
| (I) errorscore | (J) errorscore | Mean Difference (I-J) | Std. Error | Sig.a | 95% Confidence Interval for Differencea | |
| Lower Bound | Upper Bound |
| 1 | 2 | -9.900\* | .323 | .000 | -10.560 | -9.240 |
| 3 | -19.700\* | .586 | .000 | -20.898 | -18.502 |
| 4 | -30.133\* | .371 | .000 | -30.891 | -29.375 |
| 2 | 1 | 9.900\* | .323 | .000 | 9.240 | 10.560 |
| 3 | -9.800\* | .492 | .000 | -10.807 | -8.793 |
| 4 | -20.233\* | .317 | .000 | -20.881 | -19.585 |
| 3 | 1 | 19.700\* | .586 | .000 | 18.502 | 20.898 |
| 2 | 9.800\* | .492 | .000 | 8.793 | 10.807 |
| 4 | -10.433\* | .467 | .000 | -11.388 | -9.479 |
| 4 | 1 | 30.133\* | .371 | .000 | 29.375 | 30.891 |
| 2 | 20.233\* | .317 | .000 | 19.585 | 20.881 |
| 3 | 10.433\* | .467 | .000 | 9.479 | 11.388 |
| Based on estimated marginal means | | | | | | |
| \*. The mean difference is significant at the .05 level.  a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments). | | | | | | |

When the levels of a within-subjects factor represent equally spaced quantitative scores, a polynomial contrast is appropriate to use. The results for the polynomial contrasts are shown in Figure 1.5. There is a significant linear effect F (1, 29) = 5450, p <.01. High order polynomial contrasts, which include the quadratic and cubic, were also not significant. If one were to inspect the means shown in figure 1.1, you would notice a notable increase of scores over the 18 month period of treatment. A boxplot, shown in figure 1.6, visually displays the results in order to show the distributions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Figure 1.5**  **Tests of Within-Subjects Contrasts** | | | | | | | |
| Measure:MEASURE\_1 | | | | | | | |
| Source | errorscore | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
| errorscore | Linear | 15060.060 | 1 | 15060.060 | 5449.735 | .000 | .995 |
| Quadratic | 2.133 | 1 | 2.133 | .738 | .397 | .025 |
| Cubic | .807 | 1 | .807 | .275 | .604 | .009 |
| Error(errorscore) | Linear | 80.140 | 29 | 2.763 |  |  |  |
| Quadratic | 83.867 | 29 | 2.892 |  |  |  |
| Cubic | 84.993 | 29 | 2.931 |  |  |  |



Conclusions

In conclusion, there is a definite correlation between the video gameplay and the improvement of scores on the Barron Suturing Test. The scores increased by nine points each time over the 18 month period, with only minimal deviations in the scores. Therefore, the null hypothesis is rejected; there is a correlation between gameplay in three different video games will affect the error rate in the Barron Suturing Test.

The results of this study will certainly be useful to the surgical community in a variety of ways. While fine motor skills can be improved in a variety of ways, gameplay can be a fun and creative teaching method in order to improve laparoscopic surgeons in the operating room.

There are a few limitations that the study does not address; while the scores are calculated using a strict scale, the type of error is not indicated. Not all errors are fatal, nor can they be effectively compared for this study. Also, no other outside training was taken into account for the results of this study. Future studies might cross reference types of errors. They also might want to cover the amount of outside training of fine motor skills in the study in order to give a clear view of the results.

**References**

Rosser, J. C., Lynch, P. J. Haskamp, L., Gentile, D. A., & Yalif, A. (2007). The impact of video games in surgical training. *Archives of Surgery, 142*, 181-186.