COLUMBIA UNIVERSITY

Columbia University, Spring 2021

Midterm Report
Parson's Programming Puzzles: Optimizing Efficiency and Investigating the Effects of Feedback

Further research on Social Addictive Gameful Engineering (SAGE) design and computational thinking (CT)

Alexander Liebeskind | al3853

Advisors: Jeff Bender, Gail E. Kaiser

1 Abstract

This midterm report briefly summarizes research progress over the first half of the Spring 2021 semester, following upon the goals of the original research proposal and presentation (found at https://github.com/cu-sage/Documents/tree/master/2021_1_Spring). This includes a description of setup and learning research methods, a preliminary results section, a discussion of significance, and an outline of ongoing goals.

2 Progress Summary

The research proposal presented at the start of the semester was used for general guidance on early objectives. While meeting with Sandy several times to discuss data preprocessing and analysis methodology, I began research by learning and exploring SPSS through a remotely hosted Columbia computer. Preprocessing was centered around preparing data for normality testing and analysis techniques. Normality testing conducted using Shapiro-Wilk, as well as Kolmogorov-Smirnov between groups, and visualization using Normal Q-Q plots and stem-and-leaf plots. Analysis was then performed using ANOVA and post hoc tests for multiple comparisons for normal or normalizable data, and Kruskal-Wallis for a nonparametric alternative where ANOVA constraints were not satisfied. Based on early analysis outputs, the data were iteratively preprocessed to ensure result validity, avoid extraneous skew and kurtosis, and integrity between data and conclusions. Necessary score corrections and composite scores were also calculated prior to analysis. The results here cover training performance data and after puzzle (Survey 4) cognitive load data (intrinsic, extraneous, and germane). Later analysis aims are discussed in the Ongoing Objectives section.

3 Preliminary Results

In the interest of brevity, since results at this point are largely unfinalized and subject to update, the following examples of performance data and cognitive load analysis are presented to illustrate progress and aid in result interpretation. Results write-up is also intended to facilitate in later paper composition and submission.

3.1 Preprocessing

Preprocessing was conducted manually and using Python scripts to consolidate sheets, standardize data, and calculate composite scores. For example, fs2_DB_data_v0.2 was joined to survey data for cognitive load scoring by condition, since fs2_DB_data_v0.2 includes condition data. Additional data manipulation was also performed to match formatting for SPSS analysis inputs.

3.2 Normality Testing

Since analysis methodology was heavily reliant on normality testing, these tests were performed first. Shapiro-Wilk yielded a significance coefficient of 0.066 for overall cognitive load, which qualified as normal (>0.05). Training performance data yielded a Shapiro-Wilk signficance coefficient of 0.001, but negligible skew or kurtosis was detected between groups (https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php).

3.3 Performance Analysis

One-way ANOVA analysis of training performance data yielded an overall significance coefficient of $< .001$, as illustrated in Figure 3.1, indicating significant differences between conditions (< 0.05). Though this finding is not immediately relevant in and of itself, it does suggest that the data include individual condition differences and demand further review. Full multiple comparisons results are included in Figure 3.2, which depicts the condition by condition mean difference, significance, and 95% confidence interval. As an example of interpreting these data, the first section of the table with 1 in the V1 column compares condition 1 to the remaining conditions. Mean difference indicates the average difference between test scores (I-J), while Std. Error shows associated error value. Group 1 here outperformed group 2 by 69.353 on average ($p < 0.001$). Significant results (where Sig. < 0.05) in the remainder of the table are marked with an asterisk (*) on the mean difference column. 69.353 here represents raw score difference, which is determined by the puzzle scoring system.

Figure 3.1: ANOVA Summary

ANOVA

V7

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15218146.726	8	1902268.341	35.618	.000
Within Groups	172561064.970	3231	53407.943		
Total	187779211.696	3239			

3.4 Cognitive Load Analysis

Cognitive load data revealed no significant results with respect to overall cognitive load at the Survey 4 after puzzle stage. These results are summarized in Figure 3.3. The full multiple comparisons table is omitted due to the length of the output since the results are not significant. The lack of difference between condition in cognitive load perhaps implies that training difficulty is uniform across groups. Since training performance differences have appeared between groups, and thus may appear in transfer data between groups, this lack of significant disparity is likely indicative of potential differences that may arise in efficiency analysis.

4 Discussion

The above results illustrate the analysis methods and outputs generated by SPSS. It must be noted that conclusions should not yet be drawn on the basis of these outputs alone, since the comprehensive analysis has yet to be completed and the implications of significance coefficients are not finalized. However, these preliminary results nonetheless lend insight into the progress of SAGE research to date. There are also numerous factors under discussion that will guide future analysis steps (ex. binary vs. non-binary (percentage) puzzle scoring, prioritization of cognitive load significance at the transfer stage as compared to the training stage, etc.).

5 Ongoing Objectives

After finishing training performance and cognitive load analysis, the next steps are transfer performance, efficiency, and motivation. These portions require results from the performance and cognitive load stages. It should be noted that training performance

Figure 3.2: Multiple Comparisons (using ANOVA) results

Dependent Variable: V7

Multiple Comparisons

Tukey HSD

(I) V1	(J) V1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	69.353*	15.626	.000	20.85	117.85
	3	-.478	16.431	1.000	-51.48	50.52
	4	73.922*	16.084	.000	24.00	123.84
	5	1.192	16.375	1.000	-49.63	52.02
	6	95.876*	15.668	.000	47.24	144.51
	7	60.757*	17.621	.017	6.06	115.45
	8	103.263*	16.084	.000	53.34	153.18
	9	-148.065*	17.360	.000	-201.95	-94.18
2	1	-69.353*	15.626	.000	-117.85	-20.85
	3	-69.831*	17.036	.001	-122.71	-16.96
	4	4.569	16.701	1.000	-47.27	56.41
	5	-68.161*	16.981	.002	-120.87	-15.45
	6	26.523	16.301	.790	-24.07	77.12
	7	-8.597	18.186	1.000	-65.04	47.85
	8	33.909	16.701	.522	-17.93	85.75
	9	-217.418*	17.933	.000	-273.08	-161.76
3	1	.478	16.431	1.000	-50.52	51.48
	2	69.831*	17.036	.001	16.96	122.71
	4	74.400*	17.457	.001	20.22	128.58
	5	1.671	17.725	1.000	-53.35	56.69
	6	96.354*	17.074	.000	43.36	149.35
	7	61.235*	18.883	.033	2.63	119.84
	8	103.741*	17.457	.000	49.56	157.92
	9	-147.587*	18.639	.000	-205.44	-89.74
4	1	-73.922*	16.084	.000	-123.84	-24.00
	2	-4.569	16.701	1.000	-56.41	47.27
	3	-74.400*	17.457	.001	-128.58	-20.22
	5	-72.730*	17.404	.001	-126.75	-18.71
	6	21.954	16.741	.928	-30.01	73.91
	7	-13.165	18.581	.999	-70.84	44.51
	8	29.341	17.130	.739	-23.83	82.51
	9	-221.987*	18.333	.000	-278.89	-165.08
5	1	-1.192	16.375	1.000	-52.02	49.63
	2	68.161*	16.981	.002	15.45	120.87
	3	-1.671	17.725	1.000	-56.69	53.35
	4	72.730*	17.404	.001	18.71	126.75
	6	94.683*	17.020	.000	41.86	147.51
	7	59.564*	18.834	.042	1.11	118.02
	8	102.070*	17.404	.000	48.05	156.09
	9	-149.258*	18.589	.000	-206.96	-91.56
6	1	-95.876*	15.668	.000	-144.51	-47.24
	2	-26.523	16.301	.790	-77.12	24.07
	3	-96.354*	17.074	.000	-149.35	-43.36
	4	-21.954	16.741	.928	-73.91	30.01
	5	-94.683*	17.020	.000	-147.51	-41.86
	7	-35.119	18.223	.595	-91.68	21.44
	8	7.387	16.741	1.000	-44.57	59.35
	9	-243.941*	17.970	.000	-299.72	-188.17
7	1	-60.757*	17.621	.017	-115.45	-6.06
	2	8.597	18.186	1.000	-47.85	65.04
	3	-61.235*	18.883	.033	-119.84	-2.63
	4	13.165	18.581	.999	-44.51	70.84
	5	-59.564*	18.834	.042	-118.02	-1.11
	6	35.119	18.223	.595	-21.44	91.68
	8	42.506	18.581	.350	-15.17	100.18
	9	-208.822*	19.696	.000	-269.96	-147.69
8	1	-103.263*	16.084	.000	-153.18	-53.34
	2	-33.909	16.701	.522	-85.75	17.93
	3	-103.741*	17.457	.000	-157.92	-49.56
	4	-29.341	17.130	.739	-82.51	23.83
	5	-102.070*	17.404	.000	-156.09	-48.05
	6	-7.387	16.741	1.000	-59.35	44.57
	7	-42.506	18.581	.350	-100.18	15.17
	9	-251.328*	18.333	.000	-308.23	-194.42
9	1	148.065*	17.360	.000	94.18	201.95
	2	217.418*	17.933	.000	161.76	273.08
	3	147.587*	18.639	.000	89.74	205.44
	4	221.987*	18.333	.000	165.08	278.89
	5	149.258*	18.589	.000	91.56	206.96
	6	243.941*	17.970	.000	188.17	299.72
	7	208.822*	19.696	.000	147.69	269.96
	8	251.328*	18.333	.000	194.42	308.23

Figure 3.3: Multiple Comparisons (using ANOVA) results

ANOVA

Overall Cognitive Load

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	41.233	8	5.154	1.797	.075
Within Groups	1628.723	568	2.867		
Total	1669.956	576			

is distinct from transfer performance, and the current results pertain to training performance data. Overall efficiency includes instructional and performance efficiency, while motivation is calculated based upon Survey 1 - Background Information and Survey 7 - Intrinsic Motivation and TEQ portion scoring (IMFull: https://drive.google.com/file/d/1abKBVJH2POSXmLqqXNkm4C0CmqCmdTBp/view). These data will be used to determine quantitative within-subject scoring. AMOS analysis is also a long term goal, though the Columbia technology department has confirmed that AMOS is not supported through Columbia SPSS licensing. Based on reviewer comments on the fs1 paper, which included strong support of the background section, Jeff and I have also discussed making fs2 slightly more focused to enhance clarity. This includes reserving more space for intervention and result description in fs2. The primary objective remains writing up the work as a paper for submission.