Elevate Effectiveness Study

Principal Author

Dana Nakano, PhD[1]
Assistant Professor of Sociology
California State University, Stanislaus

October 2015

[1] Dr. Nakano holds a doctorate in sociology from the University of California, Irvine. He was an independent analyst for the project, which was carried out under the direction of the Elevate educational content team with the assistance of Nichols Research, Inc.

Executive Summary

Overall, the Elevate treatment group, which had access to Elevate games and training exercises, scored higher and improved more than the control group, which did not have access to the games and exercises:

- The difference in pre- and post-test scores shows a large and statistically significant improvement on the part of the Elevate treatment group. The Elevate group improved 69% more than the control group.

- Within the Elevate treatment group, people who completed an average of 4 or more Elevate training sessions per week did 17.5% better than those averaging less than 2 sessions per week and 9.5% better than those averaging 2 to 3 sessions per week

Methodology

Control and treatment groups were recruited in the spring of 2015 to test the effectiveness of Elevate over a four-week period. To be considered for either group, participants had to match all of the following criteria: reside in the United States, be fluent in English, own a smart device, be 18 years of age or older, and possess at least a high school diploma and no degree greater than a master's. Both groups were sent equivalent pre- and post-tests four weeks apart that consisted of 33 questions covering grammar, writing, listening, and math. The tests were created by Elevate independent of Elevate games and written to assess the practical skills being taught within the app. Pre- and post-test questions were written and tested in tandem so that the difficulty of the pre- and post-tests matched.

Elevate contracted a survey research company, Nichols Research, Inc., to recruit a control group of 125 individuals. Individuals in the control group were compensated $40 if they completed both the pre- and post-test. The survey for both the control and treatment groups was programmed, hosted, and tabulated by Nichols Research, Inc.

The Elevate treatment group was recruited among those who downloaded Elevate and registered an account. Within 24 hours of downloading Elevate, randomly selected individuals were emailed and asked if they would be willing to participate in a research project. Individuals meeting the above criteria were provided one free month of Elevate Pro and instructed to play

each of four different games (one in each Skill Group except Reading: Brevity, Measuring, Precision, and Retention) at least five times per week. No instructions were provided about how often to complete a daily Elevate training session. Individuals who completed the assignment received a free year of Elevate Pro. Roughly 300 individuals qualified and took the pre-test; 188 of these completed the required tasks during the four weeks; and 146 completed the post-test.

Statistical Techniques

To analyze and compare the pre- and post-test data collected from the control and Elevate treatment groups, three statistical techniques were employed to establish statistical significance: independent samples t-test, dependent samples t-test, and ordinary least squares (OLS) linear regression. Statistical significance is important because it demonstrates that the findings are reliable and that the differences observed did not occur because of random chance. The independent samples t-test is used to determine the significant difference between the means of two unique samples. In this study, this test is used to compare the total and section score outcomes of the control group and the Elevate treatment group. The dependent samples t-test allows one to determine whether the mean of two variables within the same sample differ significantly from one another. For this study, the dependent samples t-test will show if the observed differences between the pre- and post-test means within each sample group are statistically significant. While the t-tests only allow for the comparison across two variables, OLS linear regression allows one to determine the statistically significant impact of multiple variables on a single outcome variable: in this case, post-test scores. OLS regression allows one to determine how multiple variables act in concert to produce change in the mean post-test scores and tells one which variables make the most meaningful impact.

Results

Control vs. Elevate Treatment Group

Comparisons between the section and total scores for the control and Elevate treatment groups rely upon an independent samples t-test. The control and Elevate treatment groups qualify as independent samples as each set of participants is unique with no participants appearing in both groups. Figure 1 displays the mean total scores for the control group and Elevate treatment group in the pre- and post-tests. The means of each group are statistically different from each other, with the Elevate group scoring significantly higher than the control group in both the pre- and post-tests.

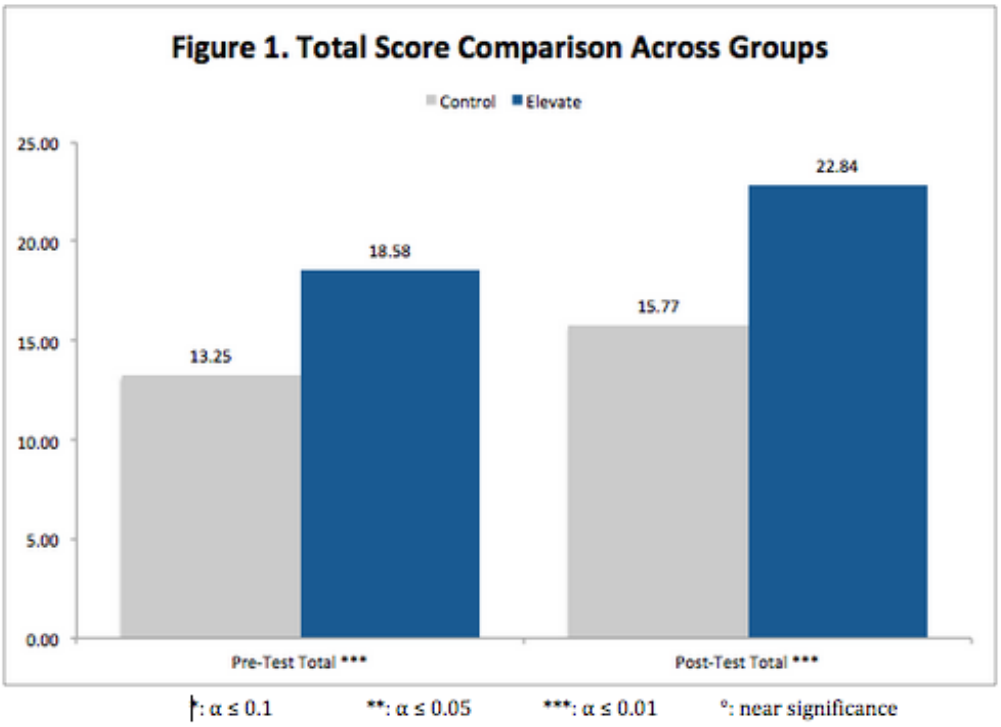


Table 1 displays the output for the independent and dependent samples t-tests. The independent samples t-test demonstrates that the comparison between the control group and Elevate treatment group in terms of mean pre- and post-test scores shows significant differences. The Elevate treatment group scored significantly higher than the control group in both the pre- and post-tests. The dependent samples t-test demonstrates that the difference between pre- and post-test scores for the control group and Elevate group, respectively, is also significant. The change in score seen from pre- to post-test is significant for both groups.

Table 1.
Independent Samples Tests

Control v. Elevate	Mean Difference
Pre-Test Total Score	4.518***
Post-Test Total Score	7.083***

Dependent Samples Tests

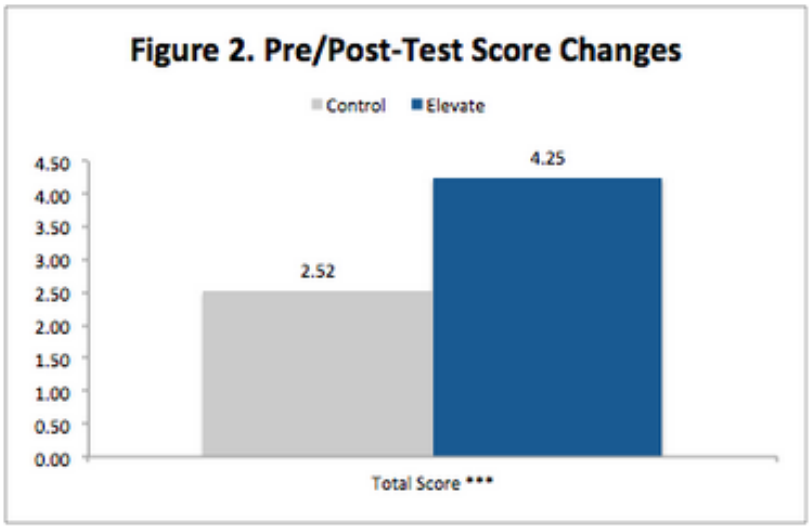
Pre-/Post-Test Total Score Comparisons within Group	N	Paired Differences	
		Mean	Std. Deviation
Control	124	-2.524***	3.908
Elevate Treatment	146	-4.253***	4.005

*: α ≤ 0.1; **: α ≤ 0.05; ***: α ≤ 0.01; °: near significance

In the pre-test, the Elevate group scored 40% higher than the control group. For the post-test, the Elevate group scored 45% higher than the control group. The higher scores persist for the Elevate group through the post-test. Table 2 in the appendix shows the difference between the mean total scores and total time

to completion of the control group and Elevate treatment group for each section and total scores in the pre- and post-test.

The Elevate group not only consistently scored higher in a statistically significant way, but it also showed greater improvement overall in comparison to the control group. Figure 2 shows the difference between pre- and post-test scores for both sample groups. The difference in pre- and post-test scores for the control group was approximately 2.5 points, an improvement of 19%. The Elevate treatment group improved by 4.25 points, a 23% improvement from pre- to post-test. The Elevate treatment group improved 69% more than the control group. These differences in improvement are statistically significant at the 0.01 level.



*: $\alpha \leq 0.1$ **: $\alpha \leq 0.05$ ***: $\alpha \leq 0.01$ °: near significance

Impact of Elevate Training Sessions on Post-Test Scores

The preceding analysis demonstrates that the group that received access to Elevate games and training did consistently better than the control group in a statistically significant way. While it can safely be assumed that the additional training positively affected the improved post-test scores within the Elevate group, this section explores how various aspects of Elevate game and training usage impacted the post-test scores of the treatment sample. Ordinary least squares linear regression is used to uncover such effects. Table 3 displays the variables that may impact the dependent/outcome variable: post-test total score. (Tables 2, 4, and 5 in the appendix provide some descriptive information for the independent variables used in the regression.) The constant reflects a catch-all variable that accounts for influences that are not captured by the included independent variables. An R^2 value indicates how much of the variance in the outcome variable, post-test total score, is explained by this particular combination of variables. The R^2 of .601 tells us that 60.1% of the variability in post-test total score can be explained by

this set of independent variables. Aside from the constant, pre-test total score, age, total number of sessions played, final difficulty level achieved for the Retention game, and difference between initial and final difficulty levels for Brevity and Measuring, all have a statistically significant impact on post-test total score.

Table 3. Ordinary Least Squares Linear Regression for Total Score

Independent Variables	Post-Test Total Score
(Constant)	***
Pre-Test Total Score	0.409***
Age	-0.133**
Average # of Sessions Played per Week	0.196**
Post-Test Completion Time	0.045
Total # of Precision Games Played	0.088
Total # of Brevity Games Played	0.070
Total # of Retention Games Played	-0.275
Total # of Measuring Games Played	-0.158
Final Difficulty Level Achieved - Precision	-0.160
Final Difficulty Level Achieved - Retention	-0.072
Final Difficulty Level Achieved - Brevity	0.525**
Final Difficulty Level Achieved - Measuring	0.297**
Difference in Initial and Final Difficulty Level - Precision	0.226
Difference in Initial and Final Difficulty Level - Retention	0.347*
Difference in Initial and Final Difficulty Level - Brevity	-0.290
Difference in Initial and Final Difficulty Level - Measuring	-0.190
N = 123 R^2 = .601	

*: α ≤ 0.1; **: α ≤ 0.05; ***: α ≤ 0.01; ***: α ≤ 0.01

It is unsurprising that pre-test scores are strongly significant in determining post-test scores. Pre- and post-test scores are positively correlated with each other; on average, for every point scored on the pre-test, the post-test score increased by .409 points. A similar strong positive relationship exists between the mean number of sessions played each week and the post-test total score. As users played more sessions, their total score on the post-test increased. However, when users played at least an average of 4 sessions per week, they saw a significant change in their post-test total score. As shown in Table 6, users who played an average of 4 or more sessions per week scored 17.5% higher than those who played less than an average of 2 sessions per week and 9.5% higher than those who played an average of 2 or 3 sessions per week.

Table 6. Impact of Sessions Played per Week on Post-Test Total Score

Average # of Sessions Played per Week	Group Mean	Difference Between Group Means		
		Less than 2	2 to 3	4 or more
Less than 2	20.5	0	1.5	3.6***
2 to 3	22	-	0	3.1**
4 or more	24.1	-	-	0
Full Sample	22.8	-	-	-

*: α ≤ 0.1; **: α ≤ 0.05; ***: α ≤ 0.01

Appendix

Table 2. Mean Scores and Times* for Control and Elevate Sample Groups

	Control Group	Elevate Group
Pre-Test Total Score	13.25	18.58
Post-Test Total Score	15.77	22.84
Pre-Test Total Time	2925.2	2644.55
Post-Test Total Time	2183.25	2132.76

*Time is in seconds

Table 4. Mean Difficulty Level Per Game Type

Difficulty Level		
		Elevate Group
Precision	Pre-Test	90.58
	Post-Test	138.93
Brevity	Pre-Test	99.85
	Post-Test	156.83
Retention	Pre-Test	59.57
	Post-Test	115.55
Measuring	Pre-Test	76.98
	Post-Test	136.14

*Difficulty ranges from 0-300

Table 5. Mean Number of Sessions Played and Games Played by Type

	Elevate Group
Number of Sessions Played	4.193
Number of Games Played	
Precision	4.936
Brevity	5.638
Retention	4.438
Measuring	5.427