Leisure Activities and Their Impacts on Cognitive Abilities

Do video games impact a child's ability to think analytically in the short term?

Venky Nagapudi, Christine Barger, Jason Baker

W241: Experiments and Causality
Dr. Micah Gell-Redman
December 17, 2019

Abstract

This study examines the impact of playing video games on the short-term cognitive abilities of teenagers. During the past fifty or so years, kids who wanted a break from schoolwork would read books, listen to music, play sports or watch television. In more recent times, however, video gaming has become the most popular leisure activity. Parents are concerned because video gaming has been linked to a number of cognitive impairments in teenagers. Our study tries to understand the causal impact of playing video games on the short-term cognitive abilities of teenagers. We conducted a field experiment consisting of a pre-test, followed by the treatment and a post-test with over 400 teenagers serving as our subjects. Our pre and post tests were short tests that served as a proxy for homework and tested mathematical, english and reasoning skills of the subjects. Participants were divided into three treatment groups and were subjected to Reading, Music or Video Games for the same time duration. We then observed whether playing video games resulted in subjects performing worse on their post-tests compared to the ones in the Reading and Music groups. Our results indicate that playing video games did result in a statistically significant negative effect on the scores.

Introduction

According to a 2019 annual report from Newzoo[1], a market research company who creates forecasts for the gaming industry, the gaming industry is set to hit $152.1 billion among a collective user base of 2.5 billion gamers globally. Projections indicate that by 2022 the gaming industry will grow to $196.0 billion, an average trend of +9.0% from 2018. More than 90% of US children play video games for more than two hours per day, which has sparked heated debates among parents and researchers as to the risks and/or benefits of gaming. A majority of the research revolves around the impact gaming has on violence and antisocial behavior, which to this day is lacking a causal link[2].

However, with gaming becoming an ever increasing leisure activity for 91% of children between two and 17[3], there has been growing concern that other maladies could be introduced such as social isolation, obesity, and cognitive impairment. It is this last issue that was of most interest

to our team. Time considerations prevented research on long-term effects; instead, we decided to narrowly focus our study on the short-term effects on mental cognition among high school students. We believe that gaming will have a short-term detrimental effect on cognition, and thus set out to study the effects of gaming compared to reading or listening to music.

Background and Past Studies

There have been multiple research projects looking into the impact of gaming on cognitive impairments with seemingly mixed results. A study[4] from Iowa State University in July 2010 indicated that “exposure to television and video games was associated with greater attention problems”, while a subsequent study from *New Directions for Child & Adolescent Development* in March 2013 suggests “that transfer from game play to more academic tasks, regardless of age, may be limited.”[5] Yet a more recent study from the *Pakistan Journal of Medical Sciences* in December 2018 finds that “Gamers exhibit better range of cognitive abilities specifically involving analogy, processing speed, deductive reasoning and mathematical intelligence...those who play video games on long-term basis, showed improvement in cognitive abilities in comparison to those who do not indulge in gaming activities.”[6]

Additional studies have indicated that short-term exposure to gaming have been linked to gaming induced physiological stress[7] and impaired concentration[8] that may lead to a decline in cognitive performance. With conflicting studies and sensationalist headlines claiming video games can cause brain damage or can alternatively augment brain power, the public gets a confusing picture on the overall effects of gaming on the brain. It is this source of confusion surrounding the short-term effects of gaming that we hope to strengthen and add to the research in this area of study.

Experiment Design

For the experiment, our subjects were teenagers between the ages of 13 and 16 years. Our experiment would involve treating some subjects with video games followed by a test of their cognitive abilities, while subjects in the control group take the cognitive test but do not play

video games. We considered several options for the control group and the design of the experiment, which can be found summarized in **Table 1**.

Table 1 *Summary of Treatment and Control Procedures for Design of Experiment*

Design	**Treatment procedure**	**Control procedure**	**Why or Why not?**
Traditional	1. Subjects play video games for 30 minutes 2. Following treatment, they take a cognitive test	Subjects are not given treatment. Instead they take the same cognitive test as the treatment group	Control group could be doing anything before the test including playing video games. We decided to specify what the control group would do for 30 minutes
Placebo Controls	1. Subjects play video games for 30 minutes 2. Following treatment, they take a cognitive test	1. Subjects would Read a book, or listen to Music for 30 minutes 2. Following treatment, they take the same cognitive test as treatment group	Treating control group with a Placebo would make the comparisons meaningful. However, how a subject fares in the cognitive test depends on not just the treatment, but also their innate abilities and other hidden variables. So we decided that we needed to eliminate this problem.
Difference in Differences with Placebo Controls	1. Subjects take a baseline test for cognitive abilities (“Pre-test”) 2. Subjects play video games for 30 minutes 3. Following treatment, they take a cognitive “Post-test” similar to “Pre-test”	1. Subjects take a baseline test for cognitive abilities (“Pre-test”) 2. Subjects listen to music or read book for 30 minutes 3. Following treatment, they take a cognitive “Post-test” similar to “Pre-test”	**We chose the D-in-D method with Placebo for our experiment since it allows us to examine the effect of treatment alone**. However, we needed to make sure that the “Pre-test” and “Post-test” were similar and that we administer all of these tests in order in one sitting.

Our Difference-in-Differences with Placebo Controls experimental design can be shown as a ROXO diagram in Figure 1.

Figure 1. ROXO Diagram of Difference-in-Differences with Placebo Controls

We divide up the group “n” into Treatment (Gaming), Control1 (Reading) and Control2 (Music) in a random fashion. Then we observe their “Pre-Test” scores for the first cognitive test. We then treat the 3 groups with the appropriate treatment, and finally observe their “Post-Test” scores for the second cognitive test.

We then analyze Pre-Test scores, Post-Test scores and the difference between the scores to understand the effect of treatment.

Design Covariates

We captured a number of details related to our subjects so that these covariates could be used for our regression analysis. These covariates included:

1. Age/Grade of subject
2. Gender
3. Subject’s geography (ex: India, California)
4. Time of test (morning, afternoon, evening)
5. Whether they own gaming consoles (could get this only for US subjects)
6. How many hours of music, reading and gaming they did (could get this only for US subjects)

Trial Study

A trial study was done with 48 subjects who fell between eighth and twelfth grade. Our team was hoping that this initial study would have sufficed, however, as will be apparent in the next few sections, due to attrition and non-compliance, our final n-value within each of our three test groups would prove too low, therefore disallowing us to run a linear regression on our results.

Experimental Procedure

Details of Test Creation

As previously mentioned, the experiment would consist of three separate sections. The first and last parts would test each subject's cognitive ability before and after treatment or control was given, whilst the middle part contained said treatment or control. The cognitive ability tests were created using questions from the Wonderlic Contemporary Cognitive Ability Test (formerly Wonderlic Personnel Test). Originally developed in 1936 by a then graduate student named Eldon F. Wonderlic, the test consisted of 50 questions which were to be answered within a twelve minute time limit. The test was created as a way to assess the learning and problem-solving aptitude of potential employees in a range of occupations and has been used by both the United States Armed Forces, as well as the National Football League, for candidate selection purposes (Wonderlic Test, 2019). The questions selected for our experiment were categorized into Math, English, and General Reasoning.

A total of 96 questions were chosen from the Wonderlic bank of test questions. Examples of specific questions can be found in *Appendix B*. For the trial study, two sets of 48 questions were randomly assigned to the pre and post experiment tests without given much thought to ensuring an equal number of math, english, and general reasoning were assigned to each. The actual test was created using the Classmarker online test maker website; further test framework details can be found in *Appendix C*.

Subject Selection and Randomization

Our team utilized personal connections, and the promise of rewards for having one of the overall group's highest score, to recruit participants into the experiment. Because of this, a pool of subjects were obtained from various geographic locations, including the San Francisco Bay Area, St. Louis, India, and parts of eastern and central Pennsylvania. Those willing to take part in the experiment were asked to fill out a short survey, which captured each participant's email address, grade level, gender, whether they owned a gaming console, and their average time spent reading, listening to music, or playing video games in a given week. Forty-eight subjects, whose grade levels ranged between eighth and twelfth, filled out the survey signaling their intent to willingly participate in the trial.

As participants completed the survey, they were placed into one of four groups based on their geographic location (PA, St. Louis, Bay Area, or India). Once the geographic groups were created, each teen from groups one through four were randomly assigned to one of the three groups, basically taking a one, two, three, one, etc. type of approach so that each of the Treatment and Control groups were made up of a similar number of teenage participants. Pictorial diagrams of the participant observation flow and ROXO randomizations can be found in figures 2 and 3, respectively.

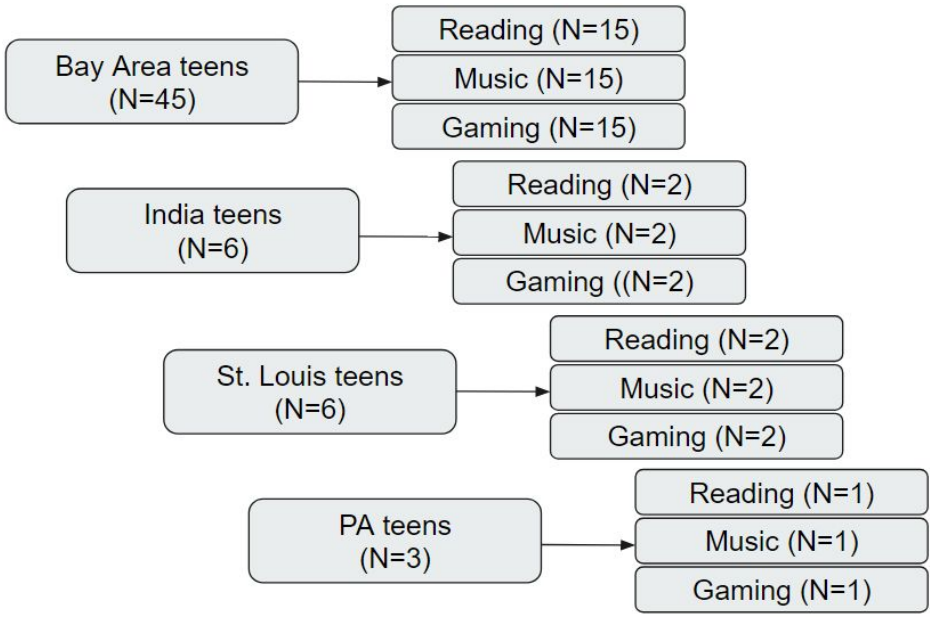


Figure 2. Observation Flow Diagram

N_BayArea	R1Treat(Gaming) R1Control1(Reading) R1Control2(Music)	O1_PreTestScore(Gaming) O1_PreTestScore(Reading) O1_PreTestScore(Music)	X1_Gaming X1_Reading X1_Music	O1_PostTestScore(Gaming) O1_PostTestScore(Reading) O1_PostTestScore(Music)
N_India	R2Treat(Gaming) R2Control1(Reading) R2Control2(Music)	O2_PreTestScore(Gaming) O2_PreTestScore(Reading) O2_PreTestScore(Music)	X2_Gaming X2_Reading X2_Music	O2_PostTestScore(Gaming) O2_PostTestScore(Reading) O2_PostTestScore(Music)
N_St_Louis	R3Treat(Gaming) R3Control1(Reading) R3Control2(Music)	O3_PreTestScore(Gaming) O3_PreTestScore(Reading) O3_PreTestScore(Music)	X3_Gaming X3_Reading X3_Music	O3_PostTestScore(Gaming) O3_PostTestScore(Reading) O3_PostTestScore(Music)
N_PA	R4Treat(Gaming) R4Control1(Reading) R4Control2(Music)	O4_PreTestScore(Gaming) O4_PreTestScore(Reading) O4_PreTestScore(Music)	X4_Gaming X4_Reading X4_Music	O4_PostTestScore(Gaming) O4_PostTestScore(Reading) O4_PostTestScore(Music)

Figure 3. Pre/Post Randomized Experimental Design (ROXO) Diagram

Experiment

The actual experiment took place over the course of several days the weekend prior to Thanksgiving. As the test administrators, we were able to limit when the test would be live based on the availability of the participants; information we asked for during the initial survey. The reason we wanted to have specific times set for the subjects to begin their testing was so that the participants would not be able to preview the test prior to their allotted test taking appointment. The actual test consisted of two ten-minute timed cognitive tests, sandwiched around an ask to complete one of the following activities for 30 minutes; playing video games, listening to music, or reading for pleasure. Because the test was online and the subjects scattered around the US and India, we were unable to control the testing environment and had to hope that most of the teens took the experiment seriously and placed themselves into an environment with as few distractions as possible. Once the tests were administered, scores for each of the pre and post tests were recorded and a score differential calculated.

As mentioned, the subjects who participated were promised prizes for those whose combined test scores were among the highest out of all the participants. Staying true to our word, we awarded prizes to the top seven scorers who complied with the entire experiment. The top scores in order of excellence, out of a total score of 96 were 83, 80, 74, 72, 68, 66, and 65, and were the results of the following tests, respectively, reading, gaming, music, reading, gaming, reading, and gaming. The top six scorers all performed better on the post test than on the pre.

Power Calculation Estimates

We utilized the one-way ANOVA power calculation, as Analysis of Variance is a commonly used approach to testing a hypothesis when dealing with two or more groups. Additionally, our hypothesis indicated a decrease in cognitive ability among those submitted to the 'Gaming' treatment; as we were concerned with a single direction difference, a One-Way test was utilized. For our trial we had 36 subjects (game_n = 12, music_n = 13, read_n = 11) with a between group variance of 0.4488 and a within group variance of 39 resulting in a power of 0.07131. **At 7% power, the trial was severely underpowered.** The most practical method for increasing power was to seek an increase sample size. Utilizing the between group variance and within group variance calculations from our trial, it was estimated that we would need 1,259 subjects across the three treatment groups. **(See details in Appendix D).** Figure 4 depicts the one-way ANOVA power calculation with the red line specifying the trial sample size at 13 per group, and the dotted blue line indicating the required sample size to achieve 80% power (420 per group).

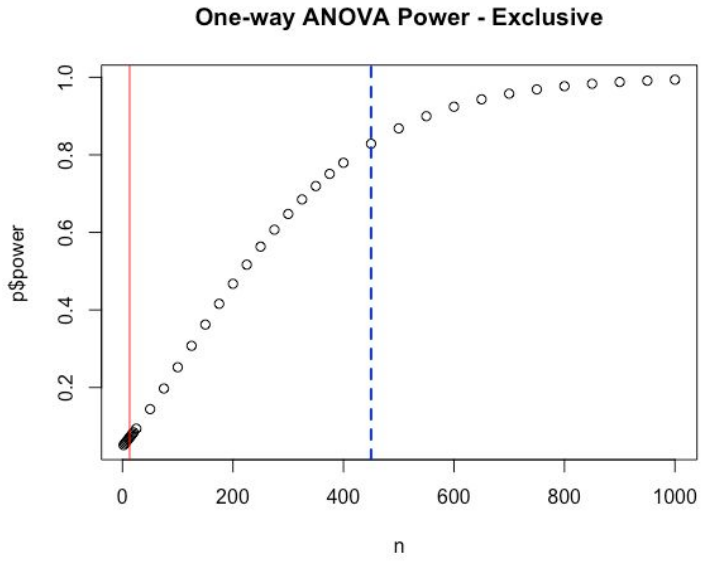


Figure 4. Graphical depiction of One-way ANOVA Power Calculation with 80% Power Indication

Analysis

As we discovered from our power calculations, our study was significantly underpowered and therefore it did not make much sense for us to run regressions on the trial study data. We

performed **ATE and CATE analysis with Randomization Inference** to confirm that we were not seeing any statistically significant results. The full results of our analysis are presented in **Appendix E**. The summary of the results are presented below in Table 2 and Figure 5, respectively.

Table 2 *Summary of Average Treatment Effect Analysis*

ATE Analysis	**All**	**Compliant**
# of Subjects	48	35
ATE	-0.76	0.14
P-value	0.38	0.51
STDEV	2.67	2.2

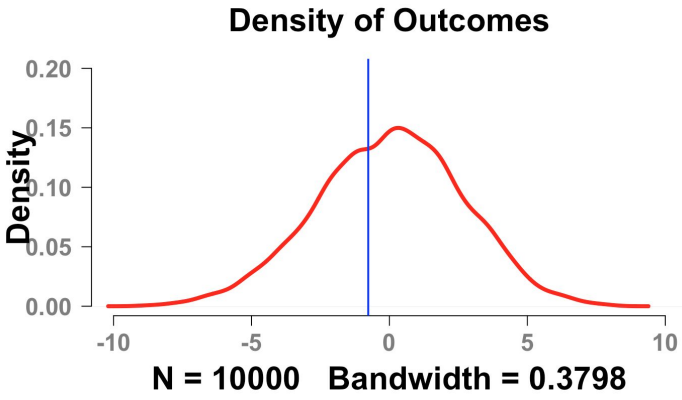


Figure 5. Density of Outcomes Using N = 10,000

We found that when we took All 48 subjects into account, there was a **negative Average Treatment Effect**, though it was not statistically significant. If we took only the 35 Compliant subjects into account, we did not observe much of an effect due to Treatment.

We also analyzed the Conditional Average Treatment Effects (CATE) attributable to the various covariates we captured. We found that being in 10th grade and owning a gaming console had a negative ATE, while being Male and taking the testing in the afternoon had a positive ATE. But

none of these were statistically significant, and we needed a much larger sample size to identify any causal effects.

Table 3 *Summary of Covariates Used in Conditional Average Treatment Effect*

Covariate	10th grade	Male	Owns Gaming Console	Afternoon Test
#Subjects	34	40	28	23
CATE	-2.3	0.89	-0.55	1.04
P-value	0.25	0.60	0.42	0.59
STDEV	3.65	2.92	3.45	3.57

Lessons learned

The results of our trial study indicated that we need to conduct a second experiment to gain meaningful causal results. We learned several lessons which were helpful in our “Production Run”.

Sample size

Our power estimates indicated that we did not get enough subjects for our trial run. To get to 80% power, we needed nearly 1250 subjects. **It was extremely difficult for us to get subjects since teenagers were averse to committing an hour of their time for the experiment**. This was very much unlike a simple survey that would only require a few minutes of their time. While it was quite challenging, we made an attempt to obtain a minimum of 400 subjects for our second experiment.

Commitment, Attrition and non-compliance

Even amongst subjects who signed up for the trial testing, we found that strong incentives failed to generate sufficient commitment to finish the testing. 13 of 48 (27%) of the subjects were either non-compliant in that they did not finish their 30 minutes of treatment, or attrited by dropping out after the Pre-Test, or were not serious about the Pre-Test or Post-Test. Since the

“carrot” of $400 in Amazon Gift Certificates failed to produce the necessary commitment, we wanted to see if the “stick” approach where the subjects took the test in a controlled school environment under strict supervision by teachers would fare better.

Pre-Test/Post-Test balance

We received feedback from some of the trial test takers that the Post-Test was a little easier than the Pre-Test. We partly expected this since the subjects were likely to fare better in the second test after they understood the format of the first test. While the Difference-in-Differences approach would account for any systematic differences between the tests, we made sure that we balanced the two tests such that there were equal number of Math, English and Logic questions.

Production Run

Experimental Procedure

For the Production Run, we searched for schools that were willing to sponsor their high-schoolers to take the test. We requested for a large number of subjects (400+) to take the test at school so that we could achieve a controlled setting. Furthermore, we asked that the testing be completed within a week. After getting rejected by several schools, we found a willing partner in **“St. John’s Public School, Vijayawada, India”**. They not only agreed to deliver 425 subjects in 9th and 10th grades, but also agreed to administer the test in their computer labs and proctor the subjects to ensure compliance to treatment and prevent attrition.

428 students took the test (140 Gaming, 115 Music, 173 Reading). However, because of a misunderstanding about the Gaming treatment, only 106 of the results could be used since a number of subjects stopped playing after one game of Pac-man and moved to the Post-Test. Furthermore, there were a number of spelling mistakes in the email that they needed to enter for each of the three tests rendering a few results unusable. Our final data set comprised of 363 subjects (106 Gaming, 100 Music and 157 Reading).

Power Calculation Estimates

We analyzed the Power first to understand if we had a sufficient number of subjects, or if further recruitment of subjects was necessary. After acquiring additional subjects for our experiment, we found our sample sizes expanding to 363 subjects (game_n = 106, read_n = 157, music_n = 100) resulted in a between group variance of 28.018, within group variance reducing to 33, and a power of 1.0 at a significance level of 0.05. **(See details in Appendix D).** Figure 6 depicts the one-way ANOVA power calculation with the red line indicating an experiment sample size at 13 per group, and the dotted blue line stipulating 80% power.

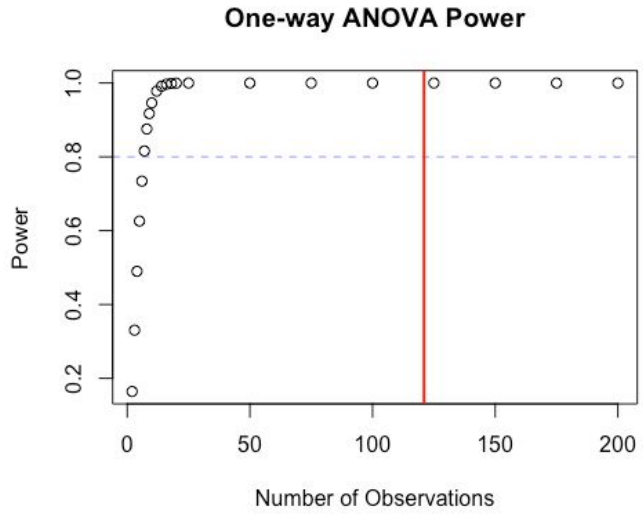


Figure 6. One-way ANOVA Power Calculation for Production Run

The power estimates indicated that our experiment was 100% powered!

Analysis

Standalone Pre-Test and Post-Test Analysis

We performed regression analysis on the Pre-Test and Post-Test scores versus Treatment Only, as well as Treatment with Covariates included. The detailed results are available in *Appendix F*. The summary of the results for these four cases is presented in Table 4.

Table 4 *Summary of Pre-Test and Post-Test Scores for Treatment Only and Treatment with Covariates*

	Dependent variable:			
	`Pre Score`		`Post Score`	
	(1)	(2)	(3)	(4)
Treated	0.313	1.131	-0.509	-0.611
	(0.584)	(0.945)	(0.556)	(0.841)
GenderMale		1.089*		0.867*
		(0.566)		(0.506)
Higher_grade		-0.081		-0.583
		(0.777)		(0.678)
`Time of Day Cat`Evening		0.537		0.655
		(0.765)		(0.680)
`Time of Day Cat`Morning		-1.396**		-0.164
		(0.684)		(0.625)
Constant	12.054***	11.642***	15.689***	15.454***
	(0.341)	(0.806)	(0.295)	(0.676)
Observations	363	363	363	363
R2	0.001	0.027	0.002	0.018
Adjusted R2	-0.002	0.013	-0.0004	0.004
Residual Std. Error	5.292 (df = 361)	5.251 (df = 357)	4.749 (df = 361)	4.737 (df = 357)
F Statistic	0.263 (df = 1; 361)	1.981* (df = 5; 357)	0.864 (df = 1; 361)	1.324 (df = 5; 357)
Note:				*p<0.1; **p<0.05; ***p<0.01

Our analysis indicated the following significant observations:

1. The Post-Test score was about 3.5 points more than the Pre-Test score. We attribute this to the fact that students become more familiar with the test format and find the second test easier. This was supported by our Trial study as well
2. Subjects who were Male fared better in both tests. This was supported by our Trial study.
3. Subjects belonging to 10th grade had a larger negative effect than students in the 9th grade. This was supported by our Trial study as well.
4. Finally, subjects who took the tests in the morning fared worse than the ones who took it in the afternoon. Again, this was seen in our Trial study.

5. In the first column, we regressed “Pre-Test scores” vs Treatment. Given that Treatment happened after Pre-Test, there should not be a statistically significant correlation, and we do not see it in this case, as expected.

Next, we performed Difference-in-Difference analysis to understand whether the difference in scores can be attributable to Treatment effects.

Difference-in-Differences analysis

We performed D-in-D regression analysis on Treatment Only and Treatment with Covariates. The details are available in *Appendix F*. The summary of the results is presented in Table 5.

Table 5 *Summary of Results for D-in-D Regression Analysis on Treatment Only and Treatment with Covariates*

	Dependent variable:	
	Score_diff	
	(1)	(2)
Treated	-0.823	-1.742*
	(0.667)	(1.051)
GenderMale		-0.222
		(0.628)
Higher_grade		-0.503
		(0.820)
`Time of Day Cat`Evening		0.118
		(0.823)
`Time of Day Cat`Morning		1.232
		(0.789)
Constant	3.634***	3.812***
	(0.360)	(0.804)
Observations	363	363
R2	0.004	0.012
Adjusted R2	0.001	-0.002
Residual Std. Error	5.756 (df = 361)	5.766 (df = 357)
F Statistic	1.534 (df = 1; 361)	0.852 (df = 5; 357)
Note:		*p<0.1; **p<0.05; ***p<0.01

We found that the treatment has a statistically significant effect on the difference in scores at a 90% confidence level.

We performed a further D-in-D regression analysis expanding the treatment group. We found that both Music and Reading groups had a higher score_diff compared to the Gaming group. The effect for the Music group is statistically significant confirming our earlier analysis. The results are shown in Table 6, below.

Table 6 *Summary of Results for D-in-D Regression Analysis Using Expanded Treatment Group*

	Dependent variable:	
	Score_diff	
	(1)	(2)
`Test Description`Music	1.009	1.794*
	(0.750)	(1.042)
`Test Description`Reading	0.705	1.563
	(0.751)	(1.321)
GenderMale		-0.210
		(0.619)
Higher_grade		-0.403
		(0.967)
`Time of Day Cat`Evening		0.180
		(0.818)
`Time of Day Cat`Morning		1.232
		(0.791)
Constant	2.811***	2.062**
	(0.561)	(0.883)
Observations	363	363
R2	0.005	0.012
Adjusted R2	-0.001	-0.005
Residual Std. Error	5.763 (df = 360)	5.774 (df = 356)
F Statistic	0.850 (df = 2; 360)	0.721 (df = 6; 356)
Note:		*p<0.1; **p<0.05; ***p<0.01

Discussion, Learnings, and Limitations

Participant Acquisition Limitations

The initial research question limited who could participate in the study in that we were specifically targeting teens who fell within the age range of a high-school student. Because of this, the task of finding participants was difficult. Our embryonic approach of seeking out teens of co-workers, teens we knew plus their friends, or attempting to contact public schools ended up being more difficult than anticipated. We found though we were asking for a single hour of the participant's time, many teens were unable to commit. In the case of asking for students within a public school, there was an added step where we relied on teachers at the school to adequately pass information of the testing on to the students in a timely manner.

Our group learned the difficulties of acquiring participants for a trial that limited the age of participants. Questions began to surface within the research team as to whether or not the experiment should be opened to all age groups, however it was ultimately decided to maintain the initial research topic and to undertake the recruitment of additional teenagers for a "production" run. We were fortunate to have found a school in Vijayawada, India willing to administer the tests to over 400 students among their ninth and tenth grade classes, providing a much larger pool of observations to be used for thorough analysis.

Length of Test

While we saw significant results with a 30 minute treatment, however it would be interesting to have conducted the experiment with subjects exposed to the treatment/control activity for a longer period of time. Part of the thinking that led to this research question was wondering what playing video games prior to a teenager attempting to do his/her homework had on their ability to satisfactorily complete their assignments.

Final Thoughts on Attrition and Non-Compliance

Even though everyone in our group has been a teenager, and some of us now parenting teenagers, we believed that by keeping the test relatively short, those who agreed to participate in the study would complete all parts of the trial. This line of thinking was reinforced during the pilot study when it was believed that awarding prizes for higher scores would result in subjects approaching the test with a moderate level of seriousness and putting for their best effort. We found, however, as was previously mentioned, approximately 27% of participants either did not comply with their treatment/control task in between the two tests, or appeared to have rapidly cycled through answers in an effort to finish each section as quickly as possible. As mentioned, this non-compliance led to a lower usable “n” value; an issue for a healthy power in addition to achieving a reliable treatment effect via regression analysis. By the same token, we were also unable to proctor the students while taking the tests; having to rely on the time spent in each test section and assume that if a teenager was in the treatment group and their time showed that they did not skip ahead to the post-test, which was indicative of the subjects having met the 30 minute treatment requirement.

In the case of the Production Run trial, participants were administered the tests in a controlled setting overseen by academic professionals, which led to a much higher rate of compliance. As mentioned previously in the experiment section, there were some spelling errors of email addresses which made it impossible to link some of the test scores, however of those for which we had good data, the rate of compliance was near 100%. We therefore bound the average treatment effect of the subgroup in which outcomes were reported regardless of whether the participants were assigned into treatment or control. This led us to the conclusion that for an experiment such as this, a controlled test taking environment is a mandatory piece.

Results Summary

The trial study was difficult to conduct based on the factors described above. Attrition and compliance accounted for ~20% of our observations which led to the inability to produce tangible results; a fact which had been made apparent through our initial power calculation, indicating the study was severely under-powered and would require a much larger sample-size

to derive meaningful insights. Furthermore, significant results were not found regarding the Average Treatment Effect when all subjects who participated in the trial study were included.

The Production Run study was successfully administered in a controlled environment on 363 subjects, 106 of which received treatment and a total of 257 received one of two controls. Power calculations revealed the study to be 100% powered. Overall, without considered covariates, receiving treatment resulted in a decrease in score between the pre and post tests of approximately one point. When covariates were included, the average treatment effect was seen with a 90% statistical significance to also cause a decrease in score, this scenario resulted in an approximate negative two point score differential. Time of day and gender were the two covariates found to have the most statistically significant impact; morning test takers scored approximately 1.5 points lower on both tests than those who took the test in the afternoon, and male subjects scored approximately one point higher on both the pre and post tests than their female counterparts. Finally, it was discovered that the ‘Listening to Music’ control group, when combined with the available covariates had a statistically significant increase in score between the pre and post test of approximately two points.

Conclusion

To conclude, the question of whether playing video games had an impact on cognitive ability was successfully answered by conducting an experiment on 363 teenagers in India. The analysis would not have been possible had it not been for the controlled testing environment, which decreased the high levels of non-compliance seen in the trial run of the experiment. The large number of participants led to the conclusion that playing video games has an average treatment effect on the outcome variable of score differential of an approximate negative one point.

Appendix A: Recruitment messaging to participants for Trial Run

Hi,
Many thanks in advance for volunteering to participate in this research study on how leisure activities influence cognitive abilities, conducted by graduate students of the School of Information and Data Science, University of California, Berkeley.

As a part of this study, you will take a couple of simple on-line tests each lasting 10 minutes. In between the two tests, you will be asked to do some reading, listen to music or play some games. The participation in this study will take less than an hour of your time in which you take the tests and do the other activities (reading, playing, listening to music). We will provide you with a couple of options for 1-hour time slots so that you can take the tests with the rest of the participants at the same time.

We are expecting around 50 students similar to you to participate in this study. And here is the exciting part! The top 3 winning scores will get some serious prizes! The highest scorer in both tests put together will get a $200 Amazon gift certificate. Two second prize winners will each get a $50 Amazon gift certificate, and four third prize winners will each get a $25 Amazon gift certificate. In the case of ties, we will use a lottery system to choose the winners. And in case you are wondering how you will get the scores, they will be available to you as a certificate as soon as you complete the tests.

To take part in this study, you just have to complete the following survey which should take less than a minute of your time.

https://www.surveymonkey.com/r/ZZ9QZCY

You will get to work on the test in the time window you specified in the survey.

We promise you that this study will be very light in terms of your effort, and will be fun as well! And you can make some serious money…not bad for just one hour of your time :-)

Look forward to hearing from you.

Thanks and regards
Christine, Jason & Venky
Graduate Students
Masters of Information and Data Science
University of California at Berkeley
Please confirm your participation in this study by emailing vnagapudi@berkeley.edu or bakerjas@berkeley.edu or cbarger@berkeley.edu before **November 8th, 2019**. We will provide you with the time slots for the tests.

Appendix B: Pre-Test and Post-Test Details

The Pre-Test and Post-Test each featured 48 questions that subjects had to answer within a strict time limit of 10 minutes. The questions were designed to test the subject's cognitive abilities. The Production Run tests included 20 Math, 15 Generic and 13 English questions each. Below are examples of some of the questions.

Question 3 of 48

There are 4 times as many used cars in the car dealership lot as there as new cars. There are 120 cars total in the dealership lot. How many new cars are in the car dealership lot?

- A) 26
- B) 27
- C) 24
- D) 25

Clear selection

Question 7 of 48

Select the word that is different from the others

- A) ample
- B) token
- C) change
- D) coin

Clear selection

Question 14 of 48

The bones that make up the spine are:

- A) femur
- B) humerus
- C) metatarsals
- D) vertebrae

Clear selection

Appendix C: Test framework with Classmarker

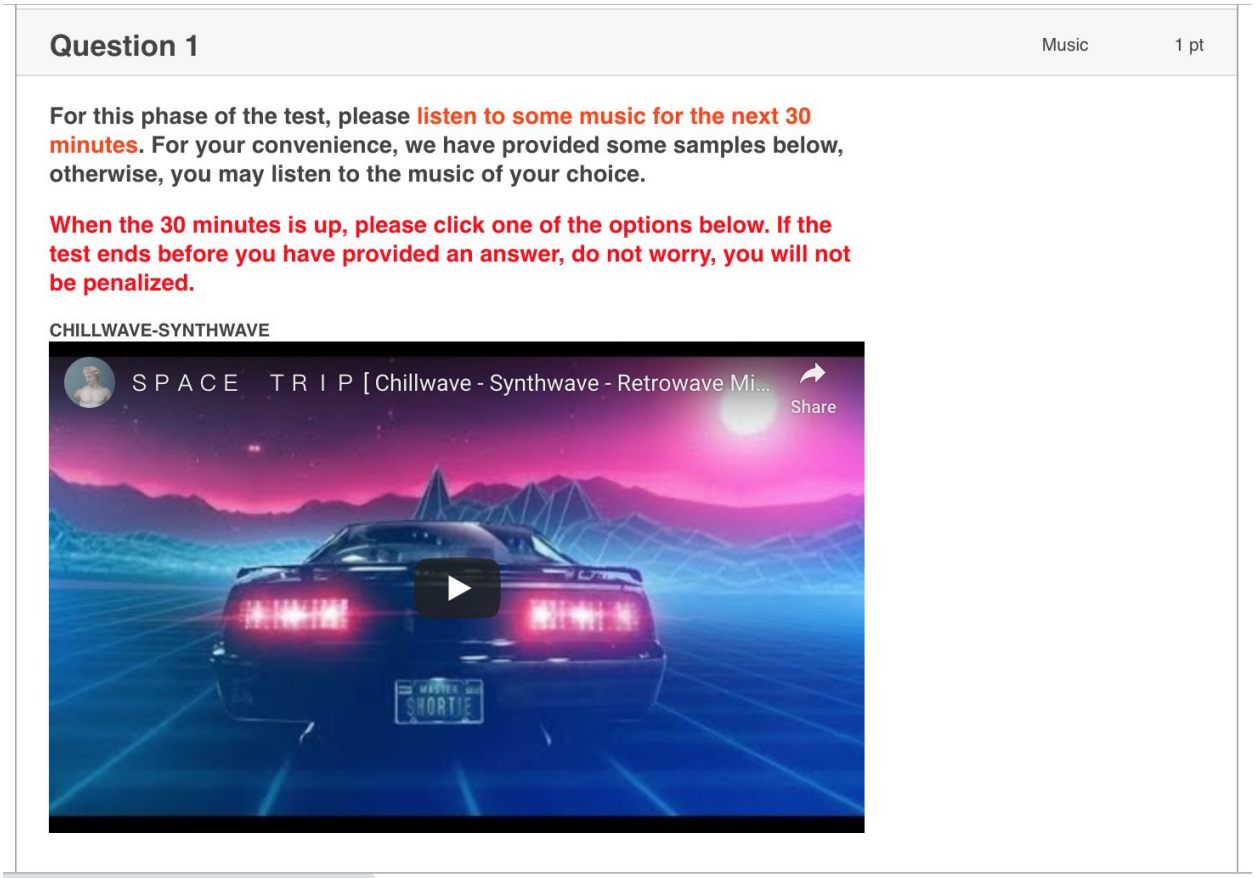
The test framework consisted of 3 tests for each of the 3 treatments: Gaming, Reading, Music. The links for the tests are below (You can click to see them):

Gaming Test

Music Test

Reading Test

Each test has a Pre-Test, a Treatment and a Post-Test. The format of the tests for Pre-Test and Post-Test were discussed in Appendix B. The format for the Treatment is shown below:



This was for the Music test. We had similar ones for Reading (on-line books) and Gaming (Pac-man).

Appendix D: Power Calculations (pdf provided separately)

Appendix E: Trial Study R Analysis (pdf provided separately)

Appendix F: Production Run R Analysis (pdf provided separately)

Appendix G: References

1. Wijman, T. (2019, June 18). *The Global Games Market Will Generate $152.1 Billion in 2019 as the U.S. Overtakes China as the Biggest Market.* Retrieved from Newzoo Web site: https://newzoo.com/insights/articles/the-global-games-market-will-generate-152-1-billion-in-2019-as-the-u-s-overtakes-china-as-the-biggest-market/

2. Ferguson, C. J. (2018, February 16). *It's time to end the debate about video games and violence.* Retrieved from The Conversation Web site: http://theconversation.com/its-time-to-end-the-debate-about-video-games-and-violence-91607

3. Reisinger, D. (2011, October 11). *91 percent of kids are gamers, research says.* Retrieved from CNet Web site: https://www.cnet.com/news/91-percent-of-kids-are-gamers-research-says/

4. Swing, E. L., Gentile, D. A., Anderson, C. A., & Walsh, D. A. (2010). Television and Video Game Exposure and the Development of Attention Problems. *Pediatrics*. Retrieved from: https://pediatrics.aappublications.org/content/early/2010/07/05/peds.2009-1508

5. Blumberg, F. C., Altschuler, E. A., Almonte, D. E., & Mileaf, M. I. (2013). The Impact of Recreational Video Game Play on Children's and Adolescents' Cognition. *New Directions for Child and Adolescent Development*, 41-50.

6. Hisam, A., & Mashhadi, S. F. (2018). Does playing video games effect cognitive abilities in Pakistani children? *Pakistan Journal of Medical Sciences*, 1507-1511.

7. Hébert, S., & Béland, R. (2005). Physiological stress response to video-game playing: the contribution of built-in music. *Life Sciences*, 2371-2380.

8. Bavelier, D., Green, C., Han, D. *et al.* Brains on video games. *Nat Rev Neurosci* 12, 763–768 (2011) doi:10.1038/nrn3135

9. *Wonderlic Test*. (2019, November 29). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Wonderlic_test