W241 Final Report:

Measuring the Effects of Positive/Negative Reinforcement on Self-Confidence with Online Surveys

Doran Bae, Brennan Borlaug, Jim Chen

Abstract

A randomized controlled trial (RCT) was conducted to expand our understanding of the effects of positive and negative reinforcement on self-confidence. Subjects, obtained via Amazon's Mechanical Turk platform, were presented an online survey containing pre-SAT mathematics questions and were asked to assess the confidence in their answers given various levels of positive and negative reinforcement. Results confirmed that the reinforcements had a significant effect on self-reported confidence levels. The effect was much larger for those who received positive reinforcement compared to negative reinforcement.

I. Introduction

Background

Self-confidence is derived from a belief in one's own abilities, skills, and expertise. It is not an innate quality, steps can be taken to improve a person's self-confidence. Doing so provides a number of perceived benefits, including a greater sense of self worth, increased happiness, enjoyment, personal fulfillment, and freedom from social anxiety/overwhelming stress. There is great value to boosting a person's self-confidence. It has the potential to make them more productive in their daily lives as well as benefit their short-term mental health. Engaging in confidence building activities and promoting self-confidence, when warranted, could provide benefits by fostering an environment that is conducive for mental/emotional growth.

One strategy for enhancing self-confidence is through positive reinforcement. This takes the form of any praise or encouragement which indicates approval in a person's actions. It is simple (and free) to provide, yet has the potential to yield higher productivities and benefit a person's emotional well being.

This research attempts to expand our understanding on the effects of positive and negative reinforcement on self-confidence by conducting a randomized controlled trial (RCT) to test whether reinforcements have a discernible effect on a person's self-reported confidence level. Online surveys served as the method for administering treatment and "level of peer agreement" provided the reinforcement (treatment). The outcome variable was a self-confidence assessment for the subject's answers to specific mathematics questions.

Hypotheses

Prior to conducting the experiment, a number of hypotheses were identified:

1) Positive/negative reinforcement will increase/decrease reported confidence levels, respectively, when compared to no reinforcement.
2) Subjects who answer questions correctly will be less affected by reinforcements.
3) The effect of reinforcement will be larger for subjects who rated their confidence levels lower in a pretreatment assessment.
4) Highly educated subjects will be less affected by reinforcements.
5) A correct response to the survey question will be predictive of self-confidence.

The experiment was designed specifically to gather evidence supporting or denying these claims.

II. Experimental Design

Surveys

Online surveys were selected as the ideal approach for conducting the study due to their relative simplicity and ability to provide a consistent, controlled experience. In the experiment, subjects were randomly assigned to complete one of two surveys at the time of their voluntary admittance into the study. Simple randomization into test groups was completed at the subject level where each subject had a 50% chance of receiving treatment. Subjects assigned to the standard survey made up the control group while those selected for the survey+reinforcement made up the treatment group. Each survey contained two primary phases: 1) Background; and 2) Test.

In the first phase of the survey, subjects were asked characteristic multiple-choice questions about themselves. This section was included to provide the information necessary to control for pre-determined covariates (and check randomization) within the study's sample. Specifically, subjects were asked about their gender, age, ethnicity, and education level. Additionally, each subject was asked to rate their general confidence in solving mathematics questions on a five-point scale: "not confident", "somewhat confident", "neutral", "mostly confident", and "absolutely confident". This provided a baseline "pre-confidence" level for future comparisons and randomization checks.

In the second phase of the survey, subjects were asked SAT-level multiple choice mathematics questions gathered from pre-SAT web resources. Upon submitting an answer, each subject was asked to rate their confidence in their answer on a five-point scale identical to the one used to measure pre-confidence level. To introduce reinforcement, the treatment group received an additional piece of information prior to assessing their own confidence - a randomized "peer agreement" statistic. The treatment took the form of a statement, "x% of respondents agreed with your answer!", where x was a randomly generated number within predetermined reinforcement ranges: 60-70% for positive, and 10-20% for negative. Among the treatment group, block randomization was conducted to ensure an equal number of positive/negative reinforcement assignments for each subject. The control group was asked to rate their confidence in their answers without the added reinforcement statement.

The average treatment effect was measured by comparing the average difference in reported self-confidence on mathematics questions between subjects who received reinforcement (positive and negative, independently) and those who did not.

Ethical Concern

It was decided that randomized peer agreement statistics would be provided to those in the treatment group to simulate reinforcement. It was critical that these values fall within a range that would provide consistent encouragement/discouragement of a person's own response providing the intended effect without being so explicit so as to generate demand characteristics. To preserve the intended treatment, reinforcement statistics were randomly generated within an acceptable range. Subjects were led to believe that these values represented their peers' responses. This minor deception was explained to each subject in the treatment group after they had finished the survey via the following disclaimer:

"You have just participated in an experiment where respondent percentages were randomly generated. Please contact {email address} for any follow up questions/concerns you have."

Sample Acquisition

To obtain a large, mostly representative sample of the population, a Qualtrics survey was distributed over Amazon's Mechanical Turk (MTurk) platform. Admittedly, the population of MTurk workers is not a perfect representation of the greater population, however it was the most effective method identified for obtaining an unbiased sample given the time and cost constraints of the experiment. 60 Human Intelligence Tasks (HITs) were published on April 6, 2016 for the pilot and 1,100 were published April 16, 2016 for the full experiment.

Pilot Study

A pilot test was conducted prior to the full experiment in order to evaluate the cost, timing, and feasibility of the completed design. In particular, the pilot study was designed to answer the following questions:

- Was the randomization successful?
- How many mathematics questions should be asked per survey?
- How much should be paid per HIT on MTurk?
- Which mathematics questions should be asked?

- Are subjects identifying the purpose of the experiment?

To prevent demand characteristics, exposure to reinforcement was limited to those who were randomized into the treatment group. For the pilot, three separate dosages were tested: 1) A one-question survey containing a random pre-SAT question; 2) A two-question survey from a pool of three pre-SAT questions and two easier questions; and 3) A four-question survey with three pre-SAT questions and one question from a pool of two easier questions. The three variations were randomly assigned for both treatment and control groups, resulting in six potential assignments.

$0.35 was paid to the first 60 people who completed the pilot survey HIT request. This was calculated by paying out at a rate slightly above the U.S. minimum wage ($5.15/hr in Georgia[1]) for the average time expected to complete the survey. Between April 6 and April 7, 2016, 75 surveys were collected in the pilot portion of the study.

Results from the pilot revealed that the four-question survey did not provide significant additional information compared to the one- and two-question versions. However, those who were randomly assigned to take the four-question survey took significantly longer to complete the HIT than other subjects. In order to maximize the sample size that could be obtained with a limited budget, it was decided that four questions were too many to ask on the final survey. A simple one question survey was also considered, as it would allow for the largest sample-to-cost ratio, but was ultimately decided against as the value of making a within-subject comparison (positive and negative reinforcement) was too large to forgo. Thus it was decided that two questions would be asked on the final survey.

The two mathematics questions that were most similar in the pilot (both in terms of a subject's correctness and confidence) were chosen for the final survey. Results from the pilot also provided reassurance that randomization was being done effectively, subject's weren't aware of the study's purpose, and subject's would be properly compensated for their efforts on the final survey.

Final Survey

On April 16, 2016, 1,100 HITs were published on MTurk for the revised two question survey. Simple randomization was used to assign subjects to treatment or control with 50% probability each. The order in which the two questions were asked was randomized for both groups as well as the order in which the answers were presented. Additionally, within the treatment group, each subject was block randomized to receive one positive and one negative reinforcement. Surveys were completed from April 16 to April 18, 2016, when the request was filled.

III. Data

1,161 surveys were collected as a result of the request for 1,100 HITs. Of these, 149 surveys were discarded for the following reasons: 115 subjects failed to complete the survey, 2 contained apparent timer malfunctions, and 32 spent fewer than three seconds on at least one of the two questions (these subjects likely completed the survey as quickly as possible to earn the reward and could be discarded as noise). 1,012 surveys (87.1%) were deemed acceptable by these standards (Table 1).

Table 2 summarizes the demography of the test sample. Surveys were distributed proportionately across demographic groups and the project team is confident that randomization was successful.

Validation of Experimental Design

For the two questions in the survey, 32% and 46% of test subjects responded with correct answers, respectively. This was larger than expected but unconcerning as it was not the variable of interest. The first step for validation (Tables 3 and 4) compared the average confidence levels in response to each of the two questions and showed no significant differences.

Another concern was that the design would violate the no-persistence assumption (i.e. past treatments would have an effect on a subject's future outcomes). This was shown to be false via a regression model (Table 5) that combined the treatment variable and order in which treatment was received.

Lastly, there was a need to ensure that differential attrition (i.e. surveys which were discarded for reasons outlined above) wasn't biasing the results. It was discovered that educational level and ethnicity had a statistically significant impact on rates of attrition (Tables 6 and 7). A theory for this finding was not yet developed at the time of publishing and the experimenters made the decision to cautiously proceed with their analysis.

IV. Results

Results are divided into four sections for independent analysis: 1) Results from the within-subject test (i.e. only subjects who received treatment); 2) Results from the between-subject test (i.e. treatment vs. control); 3) Analysis of heterogenous treatment effects (HTEs); and 4) Large regression analysis containing all covariates.

Within-Subject Test

Exposing the treatment group to both positive and negative reinforcement allowed the experimenters to make a more accurate estimation of the ATE. However, since the treatment group did not receive a control question, the individual effects of each type of reinforcement could not be separated. Computing the difference of means yields a combined ATE of ***0.65286***. Randomization inference (n = 10,000) returns an estimated p-value of ***0*** for an effect of this size, suggesting a statistically significant treatment effect.

Between-Subject Test: Basic Model

All subsequent analysis refers to a between-subject model where data is combined from both the treatment and control groups. For all subjects, each question is treated as an individual data point. The first model, a simple regression considering only the treatment variables, suggested significant treatment effects for both positive and negative reinforcement (Table 8). Additionally, the effect of positive reinforcement (**0.46**) appears to be much stronger than that of negative reinforcement (**-0.19**).

HTE Analysis: Pre-Confidence Level, Level of Education

Surprisingly, when subjects were presented with negative reinforcement to a question that they answered correctly, it appeared to strengthened the effect of the reinforcement versus those who had gotten the question wrong (Table 9). This effect was not extremely significant but still unexpected. Results also suggested that a correct answer might weaken the effect of positive reinforcement (though this HTE was not significant), another somewhat counter-intuitive result.

Pre-confidence levels did not appear to have an impact on the size of the treatment effect (Table 10). Additionally, low education level appeared to have significant impact on the effect of negative reinforcement (Table 11). While subjects with less than high school education were less affected by negative reinforcement, there was not a consistent relation between education level and the size of the treatment effect.

Large Regression Model

Finally, the project team compared five separate regression models incorporating a different number of covariates (Tables 12 & 13). The pre-confidence level of the subjects and correctness of answers (in addition to positive and negative reinforcement) had a significant relation to confidence level. Other interesting findings include: Subjects with PhD-level education tended to have a higher confidence level; Female subjects had a slightly lower confidence level than males; and “Being Asian” was a good predictor for confidence.

V. Conclusion

This research demonstrates that peer reinforcement (in terms of overall agreement) can affect a person’s self-confidence. Specifically, it suggests that if a person believes that they are backed by their peers they will be more confident in their decision and if they are opposed by their peers they will be less confident. It was also learned that positive reinforcement appears to be more effective at boosting self-confidence than negative reinforcement is at reducing it.

Application

It is uncertain how these measured changes in self-confidence would manifest themselves in the real world, but the findings in this study provide convincing evidence of a comparative effect. Given the virtually nonexistent cost to providing reinforcement, it is a strategy that can be widely deployed for marginal control over another person's self-confidence. Potential applications include: team building, mental health/rehabilitation, advertising, teaching, and more.

Future Research

A few potential areas for future exploration have been identified to expand upon these findings. Specifically, it is suggested that the experiment be repeated for different types of questions and methods of providing reinforcement. The question set should be expanded to include more abstract questions that lack concrete solutions. Having results carry over to contexts where there is no unambiguous correct response would be compelling. Different forms of reinforcement should also be tested to discover which method is most effective.

It would be useful to run a similar experiment in a controlled setting. Such a scenario could better represent a situation that subjects may actually find themselves in and allow researchers to convert self-confidence assessment ratings into tangible measurements. Designing such an experiment would open the possibilities for additional research, such as confidence in high stakes situations, compounding reinforcements, and reinforcement from various peer pools.

Acknowledgements

The research team would like to take the opportunity to thank Alex Hughes and David Reiley with the UC Berkeley School of Information for their advice and guidance in designing and conducting this experiment.

References

1) http://www.ncsl.org/research/labor-and-employment/state-minimum-wage-chart.aspx
2) https://www.qualtrics.com/support/

VI. Tables & Figures

Table 1 - Test Data Summary

	Count	Percentage
Valid surveys	1,012	87.1%
Rejected surveys	149	12.8%
Incompletion	115	9.9%
Timer malfunction (NA value for timer)	2	0.2%
Bad survey (submitted their answers in under three seconds)	32	2.8%
Total surveys	1,161	

Table 2 - Sample Demographics

Total = 1,012		Control	Treatment
Treatment and Control		505	507
Gender	Male	273	290
	Female	232	217
Age	Under 18	0	1
	18-24	95	115
	25-34	243	226
	35-44	94	95
	45-54	44	41
	55 and older	29	29
Education	Less than high school	1	3
	High school graduate	37	38
	Some college	124	142
	2-year degree (Associate's)	49	42
	4-year degree (Bachelor's)	210	195
	Professional degree (Master's)	78	77
	Doctorate (PhD)	6	10

Ethnicity	White	390	387
	Black or African American	39	26
	American Indian or Alaska Native	4	3
	Hispanic or Latino	25	34
	Asian	38	50
	Native Hawaiian or Pacific Islander	2	2
	Others	7	5

Table 3 - Regression Model for Differences Between Questions: Control

```
Call:
lm(formula = conf ~ q1, data = df[ptreat == 0 & ntreat == 0])

Residuals:
    Min      1Q  Median      3Q     Max 
-1.8713 -1.7505  0.2495  1.2495  2.2495 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.87129    0.06970  41.193   <2e-16 ***
q1          -0.12079    0.09858  -1.225    0.221    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.566 on 1008 degrees of freedom
Multiple R-squared:  0.001487,  Adjusted R-squared:  0.0004968 
F-statistic: 1.502 on 1 and 1008 DF,  p-value: 0.2207
```

Table 4 - Regression Model for Differences Between Questions: Treatment

```
Call:
lm(formula = conf ~ q1, data = df[ptreat != 0 | ntreat != 0])

Residuals:
     Min       1Q   Median       3Q      Max 
-1.98225 -1.90533  0.09467  1.09467  2.09467 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.98225    0.06605  45.152   <2e-16 ***
q1          -0.07692    0.09341  -0.824     0.41    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.487 on 1012 degrees of freedom
Multiple R-squared:  0.0006697, Adjusted R-squared:  -0.0003178 
F-statistic: 0.6782 on 1 and 1012 DF,  p-value: 0.4104
```

Table 5 - Regression Model for Potential Effect of Order of Treatment

```
Call:
lm(formula = conf ~ first * (ptreat + ntreat), data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.2759 -1.6437  0.2079  1.2079  2.4106 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.82970    0.06724  42.082  < 2e-16 ***
first        -0.03762    0.09510  -0.396 0.692411    
ptreat        0.44616    0.11520   3.873 0.000111 ***
ntreat       -0.24027    0.11749  -2.045 0.040978 *  
first:ptreat  0.02599    0.16454   0.158 0.874512    
first:ntreat  0.09187    0.16454   0.558 0.576670    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.511 on 2018 degrees of freedom
Multiple R-squared:  0.02491,   Adjusted R-squared:  0.0225 
F-statistic: 10.31 on 5 and 2018 DF,  p-value: 8.854e-10
```

Table 6 - Test for Differential Attrition via Education Level

```
        Pearson's Chi-squared test

data:  table(d$edu, d$attr)[2:8, ]
X-squared = 16.115, df = 6, p-value = 0.01315

Warning message:
In chisq.test(table(d$edu, d$attr)[2:8, ]) :
  Chi-squared approximation may be incorrect
```

Table 7 - Test for Differential Attrition via Race/Ethnicity

```
        Fisher's Exact Test for Count Data

data:  table(d$race, d$attr)[2:8, ]
p-value = 0.02583
alternative hypothesis: two.sided
```

Table 8 - Simple Regression Model for Treatment Variables

```
Call:
lm(formula = conf ~ ptreat + ntreat, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2702 -1.6174  0.1891  1.1891  2.3826

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.81089    0.04752  59.156  < 2e-16 ***
ptreat       0.45933    0.08219   5.588  2.6e-08 ***
ntreat      -0.19353    0.08219  -2.355   0.0186 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.51 on 2021 degrees of freedom
Multiple R-squared:  0.02476,	Adjusted R-squared:  0.02379
F-statistic: 25.65 on 2 and 2021 DF,  p-value: 9.982e-12
```

Table 9 - Regression Model Exploring Potential HTEs of Correct Answers

```
Call:
lm(formula = conf ~ ptreat + ntreat + pre.conf + crt + edu +
    race + gender + crt * (ptreat + ntreat), data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3320 -0.8847  0.0892  0.9167  4.1283

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.059521   0.454540   0.131 0.895830
ptreat           0.539368   0.087733   6.148 9.45e-10 ***
ntreat          -0.052586   0.085181  -0.617 0.537078
pre.conf         0.568740   0.024676  23.048  < 2e-16 ***
crtTRUE          0.859365   0.079485  10.812  < 2e-16 ***
edu2             0.112123   0.449333   0.250 0.802975
edu3             0.415968   0.441009   0.943 0.345682
edu4             0.519707   0.447710   1.161 0.245856
edu5             0.691219   0.439971   1.571 0.116328
edu6             0.889449   0.442657   2.009 0.044635 *
edu7             1.233993   0.488757   2.525 0.011654 *
race2            0.002815   0.112717   0.025 0.980081
race3           -0.400477   0.328986  -1.217 0.223631
race4           -0.182391   0.117751  -1.549 0.121550
race5            0.328021   0.098617   3.326 0.000896 ***
race6           -0.463540   0.434690  -1.066 0.286385
race7            0.165247   0.254534   0.649 0.516274
gender2         -0.172484   0.057541  -2.998 0.002755 **
ptreat:crtTRUE -0.211314   0.135380  -1.561 0.118707
ntreat:crtTRUE -0.281010   0.137442  -2.045 0.041028 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.223 on 2004 degrees of freedom
Multiple R-squared:  0.3655,	Adjusted R-squared:  0.3595
F-statistic: 60.76 on 19 and 2004 DF,  p-value: < 2.2e-16
```

Table 10 - Regression Model Exploring Potential HTEs of Pre-Confidence Level

```
Call:
lm(formula = conf ~ ptreat + ntreat + pre.conf + crt + edu +
    race + gender + pre.conf * (ptreat + ntreat), data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2799 -0.8996  0.0867  0.9142  4.1195

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.068041   0.458901   0.148  0.88214
ptreat           0.585136   0.199097   2.939  0.00333 **
ntreat          -0.085235   0.199038  -0.428  0.66853
pre.conf         0.586024   0.034703  16.887  < 2e-16 ***
crtTRUE          0.735742   0.056620  12.994  < 2e-16 ***
edu2             0.098638   0.450029   0.219  0.82653
edu3             0.395552   0.441798   0.895  0.37072
edu4             0.500738   0.448478   1.117  0.26433
edu5             0.673648   0.440786   1.528  0.12660
edu6             0.877868   0.443468   1.980  0.04789 *
edu7             1.219824   0.489284   2.493  0.01274 *
race2            0.002901   0.112841   0.026  0.97950
race3           -0.410344   0.329422  -1.246  0.21304
race4           -0.181607   0.117908  -1.540  0.12366
race5            0.316881   0.098601   3.214  0.00133 **
race6           -0.422769   0.434724  -0.972  0.33092
race7            0.163108   0.255002   0.640  0.52248
gender2         -0.169126   0.057593  -2.937  0.00336 **
ptreat:pre.conf -0.040455   0.057285  -0.706  0.48014
ntreat:pre.conf -0.023067   0.057278  -0.403  0.68720
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.225 on 2004 degrees of freedom
Multiple R-squared:  0.3641,    Adjusted R-squared:  0.3581
F-statistic: 60.39 on 19 and 2004 DF,  p-value: < 2.2e-16
```

Table 11 - Regression Model Exploring Potential HTEs of Educational Level

```
Call:
lm(formula = conf ~ edu * (ptreat + ntreat), data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.6364 -1.5070  0.1622  1.3636  3.1622 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.0000     1.0413   3.841 0.000126 ***
edu2         -2.1622     1.0553  -2.049 0.040610 *  
edu3         -1.4919     1.0455  -1.427 0.153747    
edu4         -1.4388     1.0519  -1.368 0.171539    
edu5         -1.0024     1.0438  -0.960 0.337020    
edu6         -0.6731     1.0480  -0.642 0.520784    
edu7         -0.3333     1.1248  -0.296 0.766990    
ptreat       -0.6667     1.3444  -0.496 0.620023    
ntreat       -2.3333     1.3444  -1.736 0.082782 .  
edu2:ptreat   1.5920     1.3761   1.157 0.247465    
edu3:ptreat   1.1093     1.3533   0.820 0.412473    
edu4:ptreat   1.3673     1.3715   0.997 0.318910    
edu5:ptreat   1.0947     1.3504   0.811 0.417673    
edu6:ptreat   0.9761     1.3599   0.718 0.472986    
edu7:ptreat   0.9000     1.4849   0.606 0.544517    
edu2:ntreat   2.7587     1.3761   2.005 0.045134 *  
edu3:ntreat   2.3323     1.3533   1.723 0.084960 .  
edu4:ntreat   2.2483     1.3715   1.639 0.101316    
edu5:ntreat   1.8742     1.3504   1.388 0.165336    
edu6:ntreat   2.1103     1.3599   1.552 0.120873    
edu7:ntreat   2.8667     1.4849   1.931 0.053681 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.473 on 2003 degrees of freedom
Multiple R-squared:  0.08073,	Adjusted R-squared:  0.07156 
F-statistic: 8.796 on 20 and 2003 DF,  p-value: < 2.2e-16
```

Table 12 - Regression Output of Five Models Containing Various Covariates

	Dependent variable: conf				
	(1)	(2)	(3)	(4)	(5)
ptreat	0.459***	0.455***	0.462***	0.453***	0.460***
	(0.082)	(0.068)	(0.067)	(0.067)	(0.067)
ntreat	-0.194**	-0.156**	-0.150**	-0.161**	-0.155**
	(0.082)	(0.068)	(0.067)	(0.067)	(0.067)
pre.conf		0.610***	0.594***	0.570***	0.569***
		(0.024)	(0.024)	(0.025)	(0.025)
crt		0.799***	0.758***	0.736***	0.735***
		(0.057)	(0.056)	(0.057)	(0.057)
edu2			0.114	0.108	0.088
			(0.447)	(0.450)	(0.450)
edu3			0.407	0.407	0.384
			(0.439)	(0.441)	(0.441)
edu4			0.514	0.512	0.488
			(0.445)	(0.448)	(0.448)
edu5			0.701	0.685	0.664
			(0.438)	(0.440)	(0.440)
edu6			0.886**	0.889**	0.864*
			(0.441)	(0.443)	(0.443)
edu7			1.304***	1.222**	1.213**
			(0.486)	(0.489)	(0.489)
race2				0.001	0.014
				(0.113)	(0.113)
race3				-0.418	-0.427
				(0.329)	(0.329)
race4				-0.179	-0.164
				(0.118)	(0.118)
race5				0.317***	0.319***
				(0.099)	(0.099)
race6				-0.419	-0.407
				(0.435)	(0.434)
race7				0.170	0.167
				(0.255)	(0.255)
gender2				-0.169***	-0.171***
				(0.058)	(0.058)
qtime					-0.001
					(0.001)
cftime					-0.005
					(0.005)
Constant	2.811***	0.483***	-0.052	0.111	0.191
	(0.048)	(0.089)	(0.448)	(0.454)	(0.456)
Observations	2,024	2,024	2,024	2,024	2,024
R2	0.025	0.333	0.356	0.364	0.365
Adjusted R2	0.024	0.332	0.353	0.359	0.359
Residual Std. Error	1.510 (df = 2021)	1.249 (df = 2019)	1.229 (df = 2013)	1.224 (df = 2006)	1.224 (df = 2004)
F Statistic	25.650*** (df = 2; 2021)	252.527*** (df = 4; 2019)	111.359*** (df = 10; 2013)	67.511*** (df = 17; 2006)	60.661*** (df = 19; 2004)

Table 13 - ANOVA Comparing Five Regression Models

```
Analysis of Variance Table

Model 1: conf ~ ptreat + ntreat
Model 2: conf ~ ptreat + ntreat + pre.conf + crt
Model 3: conf ~ ptreat + ntreat + pre.conf + crt + edu
Model 4: conf ~ ptreat + ntreat + pre.conf + crt + edu + race + gender
Model 5: conf ~ ptreat + ntreat + pre.conf + crt + edu + race + gender +
    qtime + cftime
  Res.Df    RSS Df Sum of Sq        F    Pr(>F)
1   2021 4608.6
2   2019 3149.8  2   1458.85 487.2323 < 2.2e-16 ***
3   2013 3042.5  6    107.27  11.9424  3.27e-13 ***
4   2006 3005.9  7     36.63   3.4956  0.000988 ***
5   2004 3000.1  2      5.72   1.9112  0.148170
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 1 - Likert Plot of Survey Results