A Model of Academic Success:
Self-control, Gender, and Academic Performance

Ben Segel-Brown
100791531

PSCI2702A Tutorial A01
Prof. Mikhail Zherebtsov
Introduction

This research project examines the relationship between gender, self-control, and academic performance, as measured by grades. These relationships are important because they could form the basis of studies to investigate how students might increase their grades, and how such methods might be applied differently by each gender. This could be important to increasing the competitiveness of Carleton or even Canadian university graduates. For instance, if self-control and academic performance are found to be higher in women than we might want to offer more self-led independent study projects for women while finding other motivators for men, such as putting men in more direct competition with one another by instituting a class ranking system. This paper looks at the students of Carleton University as a case study to test the correlations between these variables and to determine their relationship.

Accordingly, this paper tests three hypotheses. It posits that women will demonstrate more self-control than men, that women will have higher grades than men, and gender and self-control will explain a significant amount of the variation of grades, with grades being higher for women and those with more self-control.

These relationships have been found to be significant in several studies, particularly those of Angela Duckworth of the University of Pennsylvania. Duckworth found that composite self-discipline\(^1\) scores were a better predictor of academic success than IQ,\(^2\) and that self-control was

\(^1\) The terms self-discipline and self-control are used interchangeably in the literature and no distinction is make here between their meanings here.


<http://pss.sagepub.com/content/16/12/939.long> [27 March 2011], 1.
systematically related to gender. She hypothesised the relative success of women was a result of their greater self-control. More broadly, the topic of the determinants of academic performance has been studied extensively. In these studies self-control, as well as the associated trait of self-motivation, have been found to be significant predictors of academic performance.

Both Ingo Zettler and Duckworth accept a definition of self-control as “the ability to suppress prepotent responses in the service of a higher goal.” In terms of the measures of self-control, Zettler used the Retrospective Behavioural Self-control Scale that asks participants to rate the frequency with which they performed 67 tasks indicative of self-control, which generated a comprehensive score. He claims that the results of this test produced results that are strongly correlated with other measures of self-control. Similarly, Duckworth relied on a questionnaire (the Eysenck I6 Junior Questionnaire Impulsivity subscale), brief self-reports, teacher reports, parent reports and a monetary choice questionnaire (Kirby Delay-Discounting Rate.) Duckworth found all of these measures were strongly correlated. My method is somewhat similar to Zettler’s in that it is based on participant’s reflection on past behaviour,

---

4 Ibid.
5 For instance see the Journal of Learning and Individual Differences.
8 Ibid, 2.
9 A sixth measure considered, delayed choice, was not as closely related. Duckworth, Self-Discipline, 200.
specifically the total duration for which they are willing and able to focus on studying, which is assumed to reflect and therefore measure self-control.

Research Design and Measurement Procedures

Because it was specified as a requirement of the project, this study relies on data from the PSCI2702A survey. The criterion for selection for participation in the study was that the participant had to be a Carleton University student, so this can be seen to be the population of the study.

In terms of research design, this paper is a correlational case study of select data on study habits and grades collected from male and female students who are a part of the Carleton University student population. Although less than ideal as a study design, other types of study designs were infeasible. Experimentation on this topic would have been impossible and unethical, since gender cannot be assigned at random to individuals. In addition, given the brief period of time over which data was collected, a longitudinal study could not be performed; therefore the data does not allow for comparisons over time.

Privacy concerns are minimal because the data in this study are anonymous and the variables I am using are neither overly personal nor sufficiently detailed to identify a participant. As a result, there are no major ethical concerns with this research project.

The results of this study must be treated with caution however, as the data are highly problematic. The survey was completed in different ways, data-entry errors were prolific, and potent sampling bias was evident. The survey was also exceptionally long and consequently may have failed to capture participants’ interest. As a result, generalizations cannot be made from the findings from this sample to the wider student population at Carleton University. In fact, the data
of this study are insufficient to draw any broader generalizations except perhaps as part of a meta-analysis where these weaknesses are taken into account.

Some participants read and completed the survey by hand; others may have had the options read to them and the answers recorded by the surveyor. This may have introduced some bias in the data as those reading the questions may have been tempted to spare themselves the trouble of reading all answer options by picking early options, and those hearing the questions may have been biased toward picking the latter options as they heard them most recently and they can remember them best.

This dataset relied on snowball sampling. In the absence of any funding or better means, the survey was completed by the students of the class PSCI 2702A and their friends. This introduced an array of data-entry errors and potent biases. In particular, there would undoubtedly be some non-response bias because students who weren’t able to complete, or didn’t bother to complete the assignment (and their friends) would not be represented. Further, there would be some selection bias in that students would be more likely to choose friends of a similar profile who are therefore unlikely to be representative of the wider student population. It is also quite possible some students just fabricated their data, rather than having to ask or bother their friends. All in all, these data are far inferior to data used by other researchers, such as those cited in the literature review, and would not be used were it not a requirement of this assignment.

Gender and grades were measured by participants self-reporting. Self-control was calculated as the sum of the times, as self-reported in minutes per week, that participants spent studying either alone or in groups.

The validity of self-reports of gender has never been tested, but it is reasonable to assume it is fairly accurate, especially since the survey was completed in person. In comparison, self-
reported grades are known to be problematic in that “there is greater inflation by students with lower GPAs than by students with higher GPAs.” However, without funding or gross invasions of privacy, there is no more reliable measure that could be used.

Self-control could be measured in a variety of ways based on our dataset. The literature review showed that self-reports of self-control are strongly correlated with reports from teachers and parents and even more formal measures; so it seems reasonable to assume that participants’ self-reports of study time (this study’s measure for self-control) would also not be misrepresented. Several alternate, less-desirable measurements for self-control were also considered, but an examination of the descriptive statistics and some simple calculations indicated they would have produced similar results. For instance, I considered using the cumulative minutes engaged in distraction activities, or the ratio between focused activities and distraction activities as a measure of self-control.

A possible negative consequence of using total minutes spent studying as a measure of self-control is that it is quite possible that studying is a key mechanism by which the psychological attribute of self-control is translated into the measure of academic performance. However, it is a better measure than the time spent on distraction activities and is more consistent with measures used in previous studies.

11 Duckworth, Self-Discipline, 200.
A number of cases were excluded listwise. All invalid values for gender (e.g. a gender value of “120” where 1 is female and 2\textsuperscript{12} is male) were removed. Gender scores were also recoded such that males were 0 and females were 1 to enable their use in multiple regression. For the grades variable, there was considerable evidence of confusion amongst the participants as about two-thirds of students imputed percentage scores, rather than indicating scores on a 12 point scale as requested. To correct for this, scores that were reported on a 12 point scale were converted to percentages based on a scale published in the Carleton University Student Undergraduate Calendar. While it would have been more valid to convert the percentage scores to scores on the 12 point scale, this would have made the data ordinal instead of interval, as the letter grades relate to irregularly sized sets of numbers and I would consequently be unable to apply the necessary statistical tests.

For independent study time, values above 3000 minutes per week (7 hours a day) were deemed highly unlikely and excluded, as were values which were impossible when considered in combination with the person’s group study time. Three consecutive cases were deleted because they had identical values suggesting that the interviewer (each interviewer inputted 2 or 3 studies) fabricated the responses. There was also significant confusion where people inputted hours per day, hours per week, or minutes per day instead of minutes per week for independent study time. To remove many of these cases all cases with values between 1 and 99 inclusive were removed, as these low and overly specific values seemed highly suspicious. For instance, it is highly unlikely that a participant studied precisely 9 minutes per week, and even if this were

\textsuperscript{12} This refers to the value at initial entry, I latter recoded male values to be 0.
the case, they would probably round their value. For group study time, only those values that were over 2000 minutes a week were removed (4.7 hours a day), and most of these cases were already omitted due to suspicious independent study time values. In all, 81 cases were removed representing 26.6% of the cases.

In terms of the types of the variables, gender is categorical, and grades and self-control are scale variable. Consequently, independent means t-tests will be used to compare the means of Grade and self-control by gender, and a multiple regression test will be used to test the model using gender and self-control to predict grades.

**Limitations**

The study looks only at a biased sample of Carleton University students, so the results cannot be confidently generalized to any larger group, such as all university students, without further investigation. Nor would the findings of this study be perfectly representative of the Carleton University student population, due to the aforementioned sampling biases and potential misreporting by student participants.

**Description of Variables**

The survey had a total of 304 participants. Of the 288 participants who disclosed a valid gender, 148 (51.4%) were male and 140 (48.6%) were female. Thus, approximately equal numbers of participants from each gender responded to the survey (Figure 2).

Self-control scores, the sum of each participant’s individual and group study time scores (in minutes/week) were, after correction of the data and removal of cases containing unlikely values, considered to be valid in 223 out of 288 (77.4%) cases (Figure 2). The average score was
636 minutes a week (1.5 hours/day) and the standard deviation was 475.80 (1.1 hours/day)
meaning that roughly 68% of participants studied between 0.4 hours a day (23 minutes) and 2.6
hours a day (156 minutes). However, such a statement is somewhat problematic because the
results were positively skewed with large numbers of students with values less than the mean of
636, but a significant number of exceptional students who studied more than double this value
(Figure 9.) The skew occurred in part because there were a large number of students who did not
study in groups at all (Figure 10). The smallest value for total minutes per day studied was 0
minutes, and the largest was 2520 minutes (6 hours/day), creating a range of 2520 minutes.

Grades were somewhat normally distributed, with odd spikes (for instance at the modal
value 75) that likely resulted from participants or interviewers directly converting letter grades
into their corresponding numeric grade. Thus, there were about 5 respondents with grades
between 72 and 74, but about 55 respondents with scores between 75 and 76 because B grades
were converted to 75 by most interviewers. This also distorted the range because the highest and
lowest values were generated by my conversion of invalid scores (e.g. “1” meaning F) into
numeric scores (e.g. 49). The highest value was 94 and the lowest value was 49 leaving a range
of 45 percentage points.

Analysis

Hypothesis 1: Women will demonstrate more self-control than men (Figures 1, 2, 3 and
Calculation 1).

The mean self-control value for women was slightly higher (M = 666, SE=50.3), than the
mean self-control value for men (M = 609, SE=40.2). This difference was not significant t (221)
= -.898, p > .05, so the null hypothesis that there is no difference between self-control by gender
cannot be rejected. In accordance with the insignificant result, the analysis resulted in a very small effect size (r = .06).

*Hypothesis 2: Women will have higher grades than men (Figures 1, 2, 4 and Calculation 2).*

The mean value for the grades of both men and women was 77.7 (SE\textsubscript{Males}=0.66, SE\textsubscript{Females}=0.55). This difference was not significant t (262) = .009, p > .05, so the null hypothesis that there is no difference between grades by gender cannot be rejected. In accordance with the highly insignificant result, the analysis resulted in an essentially 0 effect size (r = .0006).

*Hypothesis 3: Gender and self-control will explain a significant about of the variation in grades, with grades being higher in women and those with more self-control (Figures 5, 6, 7, 8.)*

The final multiple regression model using gender and self-control as predictor variables explained 0.8% of the variation in grades (r = .09, r\textsuperscript{2} = .008). The final formula had a b value of 76.87, a B value of -.602 for gender (-.043 standardised B), and a B value of .001 for self-control (.083 standardised Beta) giving a formula for the model of 76.87 – (0.602)*gender\textsuperscript{13} + 0.01*self-control\textsuperscript{14}. Neither the gender B (p = .543) nor the self-control B (p = .240) were significant. Furthermore the F-ratio was .834 which was not significant (p = .436).

---

\textsuperscript{13} Where 0 is male and 1, female.
\textsuperscript{14} Or minutes per week spent studying. This means that an increase of 100 minutes a week of studying (or about 15 minutes a day) translates into a 1% increase in final grade.
Discussion

The most important lesson to be learnt from this study is that surveys should be clear, focused, well-administered, and ideally, preceded by qualitative research. A large proportion of students and interviewers were confused by the phrasing of the question in minutes per week, an entirely inappropriate and unintuitive measure for most students, resulting in the removal of more than a quarter of the cases ([304-223]/304=26.6%) who gave excessively low and specific or high answers, likely due to the units or math errors. This was compounded by the placement of these questions near the end of the survey, and the poor quality of the survey that suggested a complete disregard for the participants’ time as it was long, unfocused and contained repeated questions. As a result, participants likely put in even less effort to get their math right. There were also many errors that were clearly the fault of incompetent interviewers, such as the inputting of 120 for gender. Furthermore, it was discovered through discussions with fellow students that while in my head studying was an activity that demonstrated self-control, in many cases people found themselves perpetually distracted while studying suggesting that some measure of the quality of study should be integrated into the model. When cleaning the data, I exercised considerably more restraint than I would have otherwise because values that I found unbelievable (ex. > 3 hours a day of group study time) were considered normal by some of my classmates. More formal qualitative research would have better enabled me to clean the data with more certainty such that I was not imposing my own bias and would also have allowed me to develop a better measure of self-control.

Despite the limitations of the data and the absence of any statistical significance, the findings of this study are quite intriguing. The descriptive scores for self-control paint an interesting picture of the time that people spend studying. Upon entering Carleton, I was told that
students tend to spend an amount of time doing coursework at least equal to the length of class. Assuming a full course-load (5 course) of normal length classes (3 hours), this comes out to 2.14 hours a day \(((5 \text{classes} \times 3 \text{h/week})/(7\text{days/week}) = 2.14 \text{ hours/day.})\) Compared to the mean value found in this study \((M= 1.5 \text{ hours/day, SD}=1.1 \text{ hours/day})\), this suggests that about three quarters of participants were investing less than this minimum time in their studies (Calculation 4). The difference in means by gender \((M_{\text{female}} = 666m/week = 1.59 \text{ h/day}. \ M_{\text{male}} = 609m/week = 1.45 \text{ h/day})\) suggest that this problem may be more acute in men, although the small effect size and non-significance of this difference implies that the reasons for this problem are not rooted in genetic or culturally-created differences between the sexes.

The descriptive statistics of grades were less exciting. As one would expect, the grades would form a bell curve with the mode and mean centered around 77, a B+. However, the lack of any difference in grades by gender is a positive sign as it suggests that our education system is not biased against either gender. If this finding is consistent with future studies of the Carleton University population, it would be worth investigating how we achieved this equality given the findings in other academic literature that gender is a significant factor in academic performance.

The model created using gender and self-control to predict grades was neither significant nor provided the results expected given the literature review. However, if the relationships suggested by the model are confirmed in future studies, this model could be ground-breaking. Not only was being female a minor disadvantage under the model, the gender gap widened when self-control was added to the model, which could indicate either that there is an interaction effect (perhaps men study more effectively) or that there are decreasing returns for hours studied in terms of academic performance (because women studied more). It was also unexpected that the influence of time spent studying would be so small, an increase of 100 minutes a week of
studying (or about 15 minutes a day) translated into a 1% increase in final grade. I suspect that this effect would have been larger if I had not been forced to remove low study-time scores. While this effect is smaller than expected and is not significantly different from the mean, it has real world significance in that it means that an increase in self-control such that one studies an extra two hours a day could (assuming there is no selection bias or decreasing returns to scale) move one’s grade from a 77 (B+) to an 85 (A).

Conclusion

In sum, no significant relationships were found between gender, grades, and self-control. This was tested by comparing the means of grades and self-control for each gender using independent t-tests. Then, a model was created using gender and self-control to predict grades to see if gender differences in grades could be explained as a result of women’s greater self-control. In the end, no test produced significant results and the effect sizes of the relationship were very small.

This study relied on the 2011 PSCI 2701A survey which required extensive cleaning of the data. The absence of significant relationships despite the large sample size and the descriptive statistics produced, particularly self-control, have interesting implications that are discussed above. Given the findings of this study, several important questions are raised for future research such as: is study time actually correlated with other measures of self-control and why was there no relationship between gender and grades in this study when such differences have been found in academic literature?
Appendix

Figure 1: Independent Samples t-Test between Male and Female Means of the Independent Variables

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-control</td>
<td>Equal variances assumed</td>
<td>4.947</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
</tr>
<tr>
<td>GradeFix</td>
<td>Equal variances assumed</td>
<td>4.286</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Descriptive Statistics for Grades and Self-control by Gender

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
<th>95% Confidence Interval for Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Grade</td>
<td>132</td>
<td>77.7045</td>
<td>7.69727</td>
<td>.66996</td>
<td>76.3792</td>
</tr>
<tr>
<td>Male</td>
<td>132</td>
<td>77.6970</td>
<td>6.41556</td>
<td>.55840</td>
<td>76.5923</td>
</tr>
<tr>
<td>Female</td>
<td>132</td>
<td>77.7008</td>
<td>7.07197</td>
<td>.43525</td>
<td>76.8437</td>
</tr>
<tr>
<td>Total</td>
<td>264</td>
<td>77.7008</td>
<td>7.07197</td>
<td>.43525</td>
<td>76.8437</td>
</tr>
<tr>
<td>Self-control</td>
<td>117</td>
<td>609.0513</td>
<td>434.40423</td>
<td>40.16068</td>
<td>529.5080</td>
</tr>
<tr>
<td>Male</td>
<td>106</td>
<td>666.3679</td>
<td>518.09843</td>
<td>50.32217</td>
<td>566.5884</td>
</tr>
<tr>
<td>Female</td>
<td>106</td>
<td>636.2960</td>
<td>475.79923</td>
<td>31.86187</td>
<td>573.5055</td>
</tr>
</tbody>
</table>
Figure 3: Bar Graph of Self-control Means by Gender
Figure 4: Bar Graph of Grade means by Gender
Figure 5: Summary of models of academic performance

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
<th>Change Statistics</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.037&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.001</td>
<td>-0.004</td>
<td>7.08932</td>
<td>.001</td>
<td>.281</td>
</tr>
<tr>
<td>2</td>
<td>.090&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.008</td>
<td>-0.002</td>
<td>7.08261</td>
<td>.007</td>
<td>1.387</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Gender
b. Predictors: (Constant), Gender, Self-control
c. Dependent Variable: Grade

Figure 6: ANOVA tests of models of academic performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14.116</td>
<td>1</td>
<td>14.116</td>
<td>.281</td>
<td>.597</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>204</td>
<td>50.258</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>83.679</td>
<td>2</td>
<td>41.839</td>
<td>.834</td>
<td>.436</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>203</td>
<td>50.163</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Gender
b. Predictors: (Constant), Gender, Self-control
c. Dependent Variable: Grade
Figure 7: Coefficients of the regression models

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Lower B.</td>
<td>Upper B.</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>77.604</td>
<td>.689</td>
<td></td>
<td>112.702</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>-.524</td>
<td>.988</td>
<td>-.037</td>
<td>-.530</td>
<td>.597</td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>76.866</td>
<td>.930</td>
<td></td>
<td>82.619</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>-.602</td>
<td>.990</td>
<td>-.043</td>
<td>-.609</td>
<td>.543</td>
</tr>
<tr>
<td></td>
<td>Self-control</td>
<td>.001</td>
<td>.001</td>
<td>.083</td>
<td>1.178</td>
<td>.240</td>
</tr>
</tbody>
</table>

Figure 8: Coefficients of the Regression Model of Hypothesis three in APA format

B | SE B | β  
---|-----|---
Step 1
- Constant | 77.604 | .689 |
- Gender | -.524 | .988 | -.037 |
Step 2
- Constant | 76.866 | .930 |
- Gender | -.602 | .990 | -.043 |
- Self-control | .001 | .001 | .083 |

Note: $R^2 = .001$ for Step 1, $\Delta R^2 = .008$ for Step 2 ($p > .05$). No $\beta$ values were significant at $p < .05$. 
Figure 9: Frequency distribution of Self-control Scores

Mean = 636.3
Std. Dev. = 475.799
N = 223
Figure 10: Frequency Distribution of Group Study Times Contributing to Self-control Scores

Mean = 114.4213
Std. Dev. = 180.71559
N = 235
Figure 11: Frequency Distribution of Grades

Normal

Mean = 76.6462
Std. Dev. = 3.38765
N = 277
Figure 12: Relationship between self-control scores and grades

$R^2$ Linear = 0.006
Calculations

Calculation 1: Effect size calculation for Hypothesis 1.

\[ r = \sqrt{\frac{t^2}{t^2 + df}} \]
\[ = \sqrt{\frac{(-.898)^2}{(-.898)^2 + 221}} \]
\[ = \sqrt{\frac{0.806}{0.806 + 221}} \]
\[ = .0603 \]

\[ r^2 = 0.00363 \]

Thresholds used: small \( r < .10 \); medium \( r < .30 \); large, \( r < .50 \)
Therefore there was a very small effect size.

Calculation 2: Effect size calculation for Hypothesis 2.

\[ r = \sqrt{\frac{t^2}{t^2 + df}} \]
\[ = \sqrt{\frac{.009^2}{.009^2 + 262}} \]
\[ = 0.000556 \]

\[ r^2 = 0.0000003 \]

Thresholds used: small \( r < .10 \); medium \( r < .30 \); large, \( r < .50 \)
Therefore there was essentially no effect

Calculation 4: \( z \) –score of supposed minimum study time

\[ z = \frac{(x - x_{\text{Mean}})}{SD} \]
\[ = \frac{(2.14 - 1.5)}{1.1} \]
\[ = 0.58 \]

Meaning 72% of students studied less than this value. To account for the skew and possibility that some of the values between 1 and 99 were actually true, I will round this up to 75%
Bibliography


