The Competing Impacts of Negative Feedback on Academic Performance

Peter Maille
Eastern Oregon University

A laboratory experiment is used to test the hypothesis that accumulating negative feedback initially boosts, and then decreases, student academic performance. The experiment presented university students with a short quiz, and students who did not meet a preset standard received a negative feedback message. The students then took a second quiz. An analysis of the difference in performance between the first quiz and second quiz supported the hypothesized inverted U-shaped response of academic performance to negative feedback. Refining feedback strategies based on insights from this model could boost the academic performance of a broad swath of students.

Keywords: penalties, student learning, non-cognitive behaviors, performance

INTRODUCTION

This paper is motivated by my experience as a parent of two high school graduates, ten years of work as a university assistant/associate professor of economics and four years seated on my university's academic standards committee. The anecdotal evidence I gained from these experiences led me to question the effectiveness of the practices often taken to reverse a fall in academic performance or to boost low academic performance; namely the delivery of some form of punishment or academic sanction, referred to in this paper as "negative feedback". More specifically, I began to ask if the accumulation of repeated negative feedback interventions might actually decrease a student's academic performance. Academic performance can be measured by, for example, grade point average, academic awards, and scholarly recognition. Negative feedback can consist of practices like academic probation, detention, suspension, exclusion from co-curricular activities like clubs or athletic teams, and a failing grade.

This issue implicates many students. For example, at the end of the Fall term, 2015, 438 of approximately 3500 students, over 12%, at my home institution, were placed on academic probation. And we are not alone. Although this blurs the line between negative feedback related to academic performance and the feedback resulting from non-academic actions, Fabello et al. (2011) report that over the study period 59.6% of public-school students were suspended or expelled at least once between grades 7 and 12.

Equity may also be an issue. The US GAO (2018) found significant racial and gender disparities in the distribution of disciplinary actions. Considering gender, Jacob (2002) observes that nearly 60% of college students are women. He argues that a significant portion of this gender gap is the result of differences in non-cognitive skills. He finds that in elementary school, boys fail to advance a grade level more often than girls, are disproportionately represented in remedial classes, and have demonstrated behavioral problems at twice the rate of girls. Buddin (2014) attributes this to the handing in of late homework, disruptive behavior, and inattention. These arguments are echoed by Fortin et al. (2013) who find that non-cognitive factors, in

this case estimated by smoking and consumption of alcohol, are the second most important factor explaining the gender gap in academic achievement. Connecting discipline and academic performance, among eighth graders they find that detention and being sent to the office is the second most important predictor of the gender gap in grades of 8th and 10th grade boys. Owens (2016) finds that the greater level of behavioral problems of 4- to 5-year-old boys helps to explain the gender gap in schooling of 26- to 29-year-olds. Interestingly, she argues that behavior has a larger negative effect on boys than girls, in part because the same behavior tends to elicit a greater response, in this case in-grade retention.

Considering race and socio-economic status, between pre-kindergarten and 2nd grade, black students in Texas were almost five times more likely to receive out-of-school suspension while boys were over four times more likely to receive in- or out-of-school suspension than girls (Texans Care, 2018), a finding echoed by Fabello et al. (2011). Meanwhile, Farkas (2003) writes that minority and economically disadvantaged students are more likely to be retained in grade, and that this is often associated with lower engagement and poorer performance in later grades. What if the gender gap in academic performance was partly a result of, rather than ameliorated by, punishment that students received in their primary and secondary schooling? What if we had a basis for thinking that allowing a given student to advance in grade level despite poor performance would enhance future performance?

These questions implicate many thousands of students. With this in mind, the objective of this study is to gain insight into the performance-feedback relationship. This work differs from the studies described above in three ways. It considers the possibility that negative feedback can have competing effects on academic performance. It examines the effects of accumulated negative feedback rather than the effects of a single disciplinary action. It also applies across racial, gender, and socio-economic groups. This opens the door to a more finely textured discussion based on the characteristics of the individual student rather than those of a socio-demographic group.

UNDERPINNINGS OF THE MODEL

Behavioral Basis

An early observation from psychology provides the point of departure for this paper. Yerkes and Dodson (1908) presented mice with the task of selecting a white, rather than a black passageway, where incorrect selections resulted in an electrical shock. They were surprised to find that under some circumstances, "…this set of experiments did not prove that the rate of habit-formation increases with an increase in the strength of the electric stimulus up to the point at which the shock becomes positively injurious. Instead, an intermediate range of intensity of stimulation proved to be most favorable to the acquisition of a habit…" (p. 471).

This has gone on to become the Yerkes-Dodson Law. Which states more formally that the relationship between the level of arousal and performance on a task is such that over- or under-arousal decreases performance on the task (Cohen, 2011). This gives rise to an inverted U-shaped graphical representation where the intensity of "arousal" generates first a positive response, and then a negative response in terms of the independent variable.

Teigen (1994) provides a well-developed review of this study and its applications. He notes that the law was discovered by accident rather than derived based on theory, and its original formulation was unsystematic by today's standards. Indeed, Hancock and Warm (1989) refine this model by developing an alternative that is based on physical/psychological adaptability rather than arousal. Referring to the inverted U-shaped stimulus-response function, Hancock and Ganey (2003) argue that it is the common-sense appeal of the model, performance is low when a person is under stimulated and also when over stimulated, that is responsible for its wide use and appeal. However, Diamond (2005) states that the results have been replicated using a variety of different subjects, and Teigen (1994) acknowledges that it has been widely adapted with some authors arguing that this provides "…evidence of the robustness and generality of the law rather than as a sign of historical and conceptual confusion" (p. 542).

FIGURE 1
PROPOSED SHAPE OF THE RELATIONSHIP BETWEEN ACADEMIC PERFORMANCE AND NEGATIVE FEEDBACK

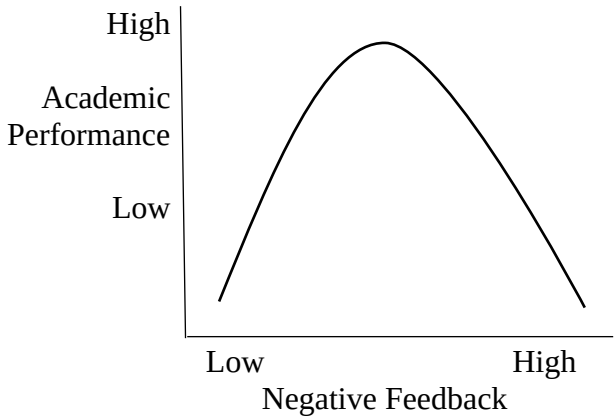


Figure 1 presents this concept in the context of the hypothesized relationship between academic performance and negative feedback examined in this paper. On the upward sloping left-hand side of the performance-feedback curve an additional unit of negative feedback will result in a gain in academic performance. On the right-hand side of the curve, the downward sloping portion, an additional unit of negative feedback results in a decrease in academic performance.

Causal Mechanisms

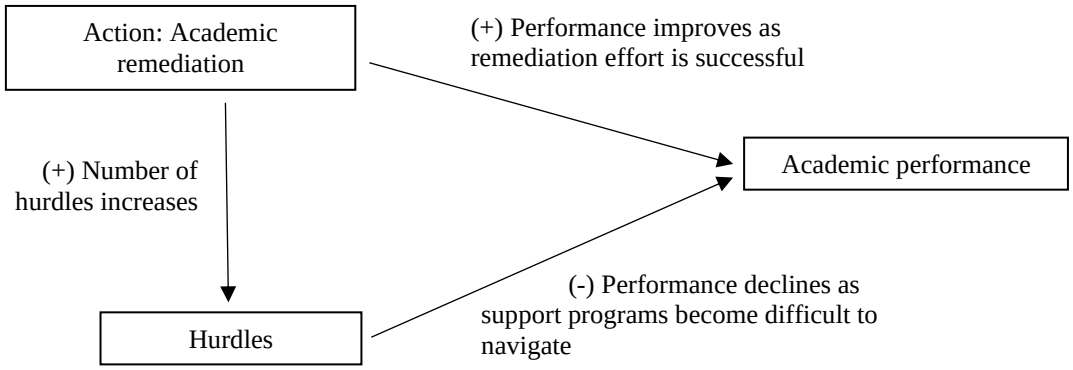
Recalling that researchers need to identify causal mechanisms (Hancock & Ganey, 2003), for illustrative purposes what follows are three possible mechanisms that may underlie the shape of the curve in Figure 1. Each mechanism includes a positive and a negative channel between some form of negative feedback and academic performance. The positive channel is the direct link between the feedback and academic performance. This is the channel that teachers and administrators use to justify delivery of the feedback. The indirect channel between the action and academic performance exerts a negative impact on academic performance. This channel may receive less attention from teachers and administrators.

In this model I assume that the negative effects of the indirect channel can overwhelm the positive effects of the direct channel as the frequency and/or intensity of the negative feedback actions accumulate, thus giving rise to the inverted U-shaped curve presented in Figure 1.

Mechanism 1

Figure 2 presents the positive and negative effects of remedial actions. Examples of these include mandatory advising sessions, required supplementary tutoring sessions, being retained in a grade, and required remedial coursework instead of electives or other courses of interest. These are not negative feedback per se but they are a direct consequence of poor academic performance. As such they could be viewed as punishment. Alternatively, remedial activities are prescribed for students who have struggled academically, and as such, are tied to the accumulation of negative feedback.

FIGURE 2
REMEDIAL ACTIVITIES AND HURDLES



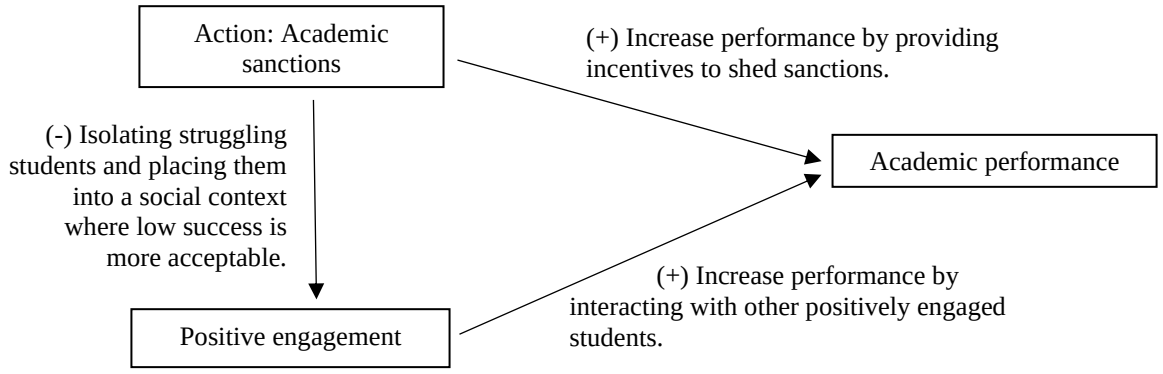
To the extent that these interventions are limited and well-delivered, students can gain from them and return to better behaviors and academic standing, thus the positive link. Considering the indirect negative channel, interventions can become impediments when they consume too much time or begin to constitute barriers in and of themselves, for example, poorly implemented remedial courses, postponement of courses of interest, or hard to schedule advising sessions. When these detrimental effects accumulate, they can begin to overwhelm the benefits of additional remedial activities, resulting in a decrease in academic performance.

Mechanism 2

Figure 3 describes one of the potential impacts of academic sanctions on academic performance. Generally speaking, sanctions in the context of Figure 3 can include, but are not limited to, failing grades, suspension, detention, loss of scholarship, and exclusion from extra-curricular activities like sports and clubs.

Sanctions are typically leveled on a student to incentivize better behavior. To the extent that this is successful, it boosts academic performance. However, sanctions can isolate a student from high-performing peers, and thereby, deny the student the benefits of positive peer relationships (see for example Lamont, 2013). They can also shunt a student away from beneficial activities. For example, rather than participating in athletic practice after school, a student who has been excluded from athletics may fill his or her time in less beneficial ways. Sanctions can also stigmatize a student, decreasing positive interactions with the student body at large. Again, the assumption is that after some level of sanction, additional sanctions could decrease academic performance because the motivation to boost academic performance is overwhelmed by the ill-effects of isolation/stigmatization/exclusion.

FIGURE 3
ACADEMIC SANCTIONS AND MARGINALIZATION



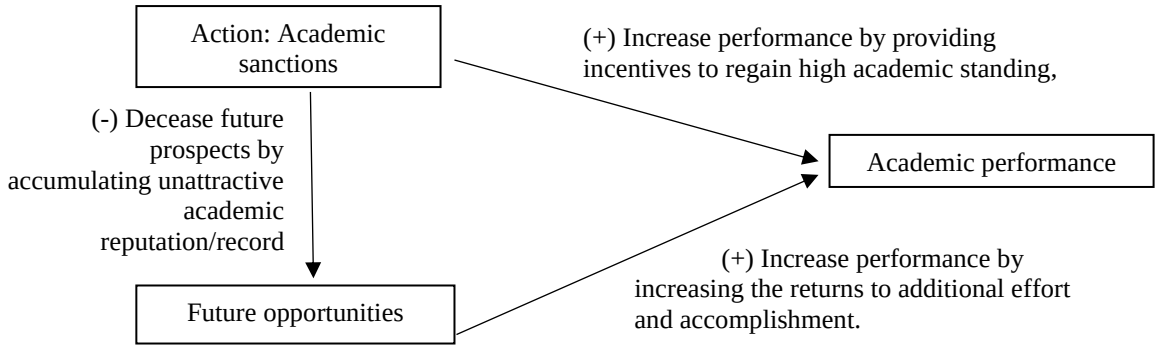
Mechanism 3

The mechanism presented in Figure 4 is also operationalized through the leveling of sanctions. In a positive sense, the desire to maintain a good record and the benefits that can flow from such a record incentivize students to boost academic performance. However, failing grades, detentions, and exclusion from athletics have important negative effects also. Specifically, mechanism 3 speaks to the decline in future opportunities that can result quite quickly from a subpar academic record.

Students will work hard to get into an Ivy League college or earn an honors diploma in high school. When that is no longer an option, which can happen as early as 9th or 10th grade, student effort decreases. In economic terms, the opportunity cost of additional decreases in academic performance declines as a student's academic standing declines.

These three mechanisms are illustrative, and all could generate the hypothetical performance-feedback curve represented in Figure 1. However, there is a more basic question, specifically, can the presence of the inverted U-shaped relationship between academic performance and negative feedback be demonstrated? That is the intent of the experiment described below.

FIGURE 4
ACADEMIC SANCTIONS AND FUTURE OPORTUNITIES



HYPOTHESIS AND EXPERIMENTAL METHODOLOGY

If Figure 1 is correct, then, all else equal, students on the upward sloping portion of the performance-feedback curve would respond positively to an additional unit of negative feedback, while students on the downward sloping portion of the curve would respond negatively. One approach to testing this hypothesis is to present students with a task, then negative feedback, and then a second task. To the extent that students can be classified based on how much negative feedback they have received prior to the experiment, the

expectation is that the performance of students positioned on the left-hand side of the curve in Figure 1 increases from the first task to the second task, while the performance of students on the righthand side of the curve decreases.

To test this, I developed a simple experiment whereby university students responded to a short survey and then took a series of two quizzes. The experiment was conducted at the university testing center where students participated under the supervision of testing center staff. The survey questions asked students to report on high school and college grade point average, presence on the Dean's List, age, gender, if they had previously taken an economics course, and if they had ever fallen into academic probation. Once these questions were answered, they were presented with multiple-choice quiz questions selected from a test bank for a 100-level introductory course in economics. Each student's performance on this first quiz provided a baseline measure of academic performance.

Depending on the results from this quiz, each student got a message saying that they either 1) did well enough to consider skipping 100-level economics, and moving directly to the 200-level classes, or 2) that their score indicates they should take the 100-level survey of economics course before taking the 200-level courses. This latter message constituted the experiment's negative feedback. In both cases, the message reporting their results on the first quiz was followed by a request to take a second quiz to confirm their results. The change in quiz scores from the first to the second quiz among students who received negative feedback, provided the basis for the experiment's test of the hypothesis.

Initially all university students were invited to participate in this experiment via a flyer that explained the experimental procedure. The flyer also offered a chance to win a $25 gift certificate as an incentive to participate. Twenty-nine students participated in the initial run of the experiment. This resulted in a sample size that was too small to generate statistically significant results. Consequently, I reran the experiment, incorporating the following observation from the first run.

Quiz 2 scores were uniformly and substantially lower than expected. Upon examination I determined that quiz 2 was more difficult than quiz 1, likely because this pool of questions included some that were more challenging. To account for this, all students quiz scores were de-meaned. Therefore, each student's quiz 1 and quiz 2 score is relative to the average of all participants for that quiz.

In the second round of the experiment students from 3 freshmen classes with the same instructor participated. Each class came in as a group and students were allowed to leave when finished. Many students were not able to respond to all the queries on portions of their academic records. The exception to this was the case of academic probation. To the extent that academic probation is the culmination of multiple episodes of negative feedback, this was used as a proxy to separate students into two groups. The group that had experience with academic probation were assumed to have received a substantial amount of negative feedback, and therefore, be positioned on the downward sloping portion of the performance-feedback curve depicted in Figure 1. They are classified as "Group B." Students having no experience with academic probation were assumed to be positioned on the upward sloping portion of this curve. They are referred to as "Group A."

Forty-five students participated in this round of the experiment with all but 1 student taking both quizzes and giving information on probation history.

RESULTS

Nineteen students received the negative feedback message based on their performance on quiz 1. Of these students, 17 had no history of academic probation (Group A), and 2 students (Group B) did. Figure 5 shows the average change in quiz scores among the students receiving the message recommending a 100-level introductory course.

FIGURE 5
AVERAGE DE-MEANED CHANGE IN QUIZ SCORE FOR STUDENTS RECEIVING NEGATIVE FEEDBACK

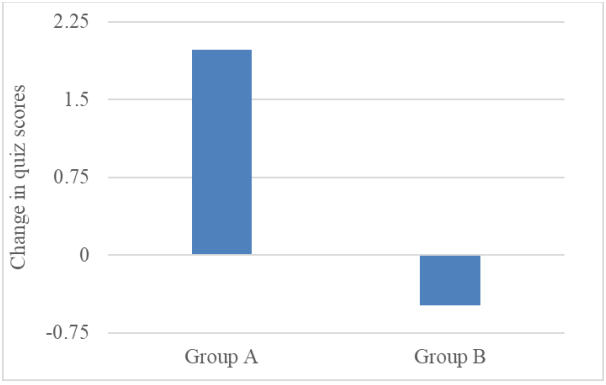


Figure 5 indicates that the students in Group A increased their quiz score an average of approximately 1.98 points relative to the quiz average, while those in Group B decreased their score by approximately 0.49 points relative to the average. To the extent that using academic probation to separate students into those who have received relatively little negative feedback and those who are likely to have received substantial negative feedback, these results are consistent with the theoretical model presented in Figure 1.

Importantly, the average scores for quiz 1 were very similar between these two groups of students. Group A students had an average 6.2, (standard deviation = 1.7) while those in Group B had an average quiz 1 score of 7 (standard deviation = 1.4). This implies that the change in quiz scores was not a function of differential performance on quiz 1.

A statistical test of these results is challenging given the small sample size. For example, it would be difficult to make the standard assumptions of normality required by least squares regression or a Students' t-test of the difference between the means. With this in mind, I used the non-parametric Mann-Whitney *U* statistic. Table 1 presents the changes in quiz scores, relative to the mean for each quiz, by group.

As Table 1 shows, aside from the challenge of having a small sample size, there are numerous ties in the data. To account for these ties, I used the mid-range approach, and the resulting rankings are in the right two columns of Table 1. These rankings were used to test the null hypothesis that there is no difference between the responses of these two groups of students, against the alternative hypothesis that students in Group A would increase their scores more than the students in Group B.

TABLE 1
SUMMARY OF EXPERIMENTAL RESULTS

Each Students Change in Quiz Scores from Lowest to Highest		Ranking of Changes in Quiz Scores	
Group A	Group B	Group A	Group B
-1.49	-1.49	18.5	18.5
-0.49		16.5	
-0.49		16.5	
0.51	0.51	13	13
0.51		13	
0.51		13	
0.51		13	
2.51		8	
2.51		8	
2.51		8	
2.51		8	
2.51		8	
3.51		4.5	
3.51		4.5	
4.51		2.5	
4.51		2.5	
5.51		1	

The calculated *U*-statistic was 5.5 which is less that the critical value of the *U*-statistic at the 90% confidence level, 6. Thus, I reject the null hypothesis and accept the alternative stating that the students with no history of academic probation increased their scores more. Given the ties in the data, this confidence level is conservative (Sokal & Rohlf, 1980).

DISCUSSION

Using academic probation to identify "high negative feedback students" these students tended to increase scores less when given additional negative feedback. Thus, the experimental results support the hypothesis that as negative feedback accumulates student response to additional feedback can switch from increasing to decreasing academic performance as depicted in Figure 1. These results raise a number of interesting issues.

One of these is how to disentangle the many forces at work in a given case. For example, even in this simple controlled setting it is difficult to classify the above described experiment in terms of any single mechanism summarized in Figures 2 through 4. A recommendation that a student take an additional course could be viewed as a "hurdle" (mechanism 1) but this would only be partly correct because anticipation of a hurdle is different from an actual hurdle. Future research could focus on elaborating and testing causal mechanisms. For example, the experiment outlined in this paper may shed light on an "expectations" mechanism. Wigfield and Eccles (2000) review the "expectancy-value theory of motivation" stating that peoples' performance is partly a direct function of their belief in how well they will do. They argue that children's belief in their abilities and likelihood of success were the strongest predictors of subsequent scores in math. This persisted even when past achievement was controlled for. Could it be that the students in Group B have acquired lower expectations independent of their abilities, and therefore, after receiving

negative feedback had a worse outcome for quiz 2? Meanwhile, their peers who had been more successful in avoiding negative feedback, when given the same message, responded by boosting their performance.

In addition, the curve presented in Figure 1 is a generalization. Diamond (2005) reviewing the original Yerkes Dodson experiment explains that the inverted U-shape of the curve depends on the difficulty of the task. Specifically, more difficult tasks have demonstrated the inverted U-shape while simple tasks are more likely to be a strictly positive function for all levels of stimulus. Where the curve actually bends, how high it is, whether it might be skewed to the left or right, are open questions. Shedding light on these questions could provide insights into how a given student may be expected to respond to negative feedback, and consequently allow us to tailor feedback to optimally impact academic performance.

This last point opens the door to an interesting issue—should negative feedback be delivered equally to all students? For example, fairness is used as a reason schools implement "zero tolerance" policies (Lamont, 2013). However, the model depicted in Figure 1 implies that what may appear fair in terms of how it is administered, could actually affect students' academic performance quite differently.

CONCLUSIONS

The empirical results reported here are based on a small sample and reach statistical significance at only the 90% level. However, the intuitive basis for inquiry into the inverted U-shaped performance-feedback relationship is strong and the stakes are significant. Therefore, research to confirm the results presented here would be worthwhile. In addition, future work into causal mechanisms like those illustrated in Figures 2 through 4 could boost our understanding student response to negative feedback, and point towards levers that may be pulled to improve academic performance. Better defining the shape of the curve in Figure 1 is also important. Knowing how this curve might shift as a function of socio-demographic and individual characteristics would provide educators with a tool they can use to boost the academic performance of their students. Research into the impact of accumulated negative feedback on academic performance could help unlock the potential for a large number of students.

ACKNOWLEDGEMENTS

I would like to thank Dr. Steve Tanner, Associate Professor of Mathematics, Eastern Oregon University, for his support in the development of this analysis, and Liz Becker for providing access to her students.

REFERENCES

Buddin, R. (2014). *Gender Gaps in High School GPA and ACT Scores*, Information Brief 2014-12. ACT Research and Policy. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2014-12.pdf

Cohen, R.A. (2011). Yerkes–Dodson Law. In: J.S. Kreutzer, J. DeLuca & B. Caplan (Eds.) *Encyclopedia of Clinical Neuropsychology* (2737–2738). New York, NY: Springer.

Diamond, D.M. (2005). Cognitive, Endocrine and Mechanistic Perspectives on Non-Linear Relationships between Arousal and Brain Function. *Nonlinearity in Biology, Toxicology, Medicine*, *3*(1), 3-7.

Fabelo, T., Thompson, M.D., Plotkin, M., Carmichael, D., Marchbanks, M.P., III., & Booth, E.A. (2011). *Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement*, Public Policy Research Institute at Texas A&M University. New York, NY: Council of State Governments Justice Center. Retrieved from https://csgjusticecenter.org/wp-content/uploads/2012/08/Breaking_Schools_Rules_Report_Final.pdf

Farkas, G. (2003). Racial Disparities and Discrimination in Education: What Do We Know, How Do We Know It, and What Do We Need to Know? *Teachers College Record, 105*, 1119–1146.

Fortin, N.M., Oreopoulos, P., & Phipps, S. (2015). Leaving Boys Behind: Gender Disparities in High Academic Achievement. *Journal of Human Resources*, *50*(3), 549-579.
Hancock, P.A., & Ganey, H.C.N. (2003). From the Inverted-U to the Extended-U: The Evolution of a Law of Psychology. *Journal of Human Performance in Extreme Environments*, *7*(1), 5-14.
Hancock, P.A., & Warm, J.S. (1989). A Dynamic Model of Stress and Sustained Attention. *Human Factors*, *31*(5), 519–537.
Jacob, B.A. (2002). Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, *21*(6), 589-598.
Lamont, J.H. (Lead Author). (2013). Out-of-school suspension and expulsion. *Pediatrics*, *131*(3), e1000-e1007.
Owens, J. (2016). Early Childhood Behavior Problems and the Gender Gap in Educational Attainment in the United States. *Sociology of Education*, *89*(3), 236–258.
Sokal, R.R., & Rohlf, F.J. (1981). *Biometry* (5th ed.). New York: W.H. Freeman and Company.
Teigen, K.H. (1994). Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology*, *4*(4), 525–547.
US GAO. (2018). *K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities*, GAO-18-258. Retrieved from https://www.gao.gov/assets/700/690828.pdf
Wigfield, A., & Eccles, J.S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.
Yerkes R.M., & Dodson J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.