The Impact of Teachers on Grades

by

Alex McLeod

An essay submitted to the Department of Economics in partial fulfillment of the requirements for the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada
August 2010
copyright © Alex McLeod 2010

Abstract

My paper studies how a teacher's design of course evaluation affects students' grades. I build a theoretical model and use numerical examples to illustrate the results. The university values the education of its students and the sorting of its students into levels of ability. The teacher values teacher evaluations and the standard deviation of the grades in his course. The university sets the teacher's valuation of the teacher evaluation relative to the standard deviation according to the university's valuation of education relative to sorting. The teacher attempts to achieve these objectives by designing a course evaluation that achieves a high class average and a high class standard deviation. The teacher is bound to grade fairly, and can only affect grades by choosing the weight on the assignment relative to the exam. When the university values education highly relative to sorting, the teacher chooses a weighting that achieves high grades and a low standard deviation. When the university values sorting highly relative to education, the teacher chooses a weighting that achieves low grades and a high standard deviation. In the model extension the teacher can also choose how difficult to make the exam relative to the assignment. When the university values sorting highly relative to education, the teacher can achieve lower grades and a higher standard deviation.

Acknowledgments

I would like to thank my supervisor Amy Sun for all her help.

I would like to thank all my friends who often discussed the paper with me and offered advice.

Table of Contents

1. Introduction

Students are the main input into the education process, and one of their strongest motivations to exert effort is to receive good grades. Students want to receive good grades because other universities and employers will use the grades as a signal of the students' ability. If the other universities and employers think a student is of high ability, then they will reward him with acceptance into their program or with higher wages. However, the grade a student receives is not completely within his control. The determination of grades is a process among students, teachers, and the university, where all the actors have their own motivations and roles. If grades are used as a signal of the ability of a student, an accurate signal would give the same impression to an employer whatever the impact the teachers and university had on the student's grades. Though schools and employers do take many factors into consideration when interpreting a grade, one factor that must be considered to maintain to value of grades as a signal of student ability is the role of teachers and the university in the grade determination process. This paper is an investigation into the likely impact of teachers and universities on grades. Specifically, if teachers are motivated to achieve a high mean and a high standard deviation, how will they set the weights and difficulties on tasks?

There are two theories explaining why people who attain higher levels of education earn higher wages. The first is the human capital theory. It posits that people who go to school gain productive skills and then are paid a greater wage because they are more productive. The second is the screening theory. It posits that people who attain higher levels of education also tend to have other characteristic that are desirable to an employer such as intelligence, problem solving abilities, perseverance, good health, stable lifestyle, etc (Wiess 1995). For a student entering university, there are two socially productive functions the university performs. First, it educates the student to improve his human capital. Second, it screens the student to signal to the labour market the ability of the student.

In some cases one objective might be more important than the other to the university. For instance, most students who graduate from undergraduate programs do not continue to higher levels of education or get asked for their grades by their employers. So sorting these graduates into precise levels of ability is not important relative to educating them. However, each of these students will have attained a degree, which is a signal to employers of their ability. Educating these students, so that the degree can be a signal of good ability is relatively much more important. In contrast, the final years of a PhD, while both educating and sorting are still important, sorting is relatively more important because the graduates from a PhD program will be moving on the positions in universities and jobs of high productivity, where the employers are very concerned with the ability of the graduate.

A university will give incentives to its teachers to perform these two functions. It will want its teachers to teach well, as well as sort students thereby enabling higher ability student to potentially separate themselves from lower ability students. The two mechanisms a university has to identify whether a teacher is properly performing these functions are teacher evaluations and whether the grades in the teacher's course are well spread out. A teacher might respond to teacher evaluations in a variety of ways, such as changing teaching practices, where he devotes his effort, or the type of tasks he requires of his students. One main consideration, and the one that is the focus of this paper, is changing the students' grades in order to have his students more satisfied with their grades and reward the teacher with better teacher evaluations. If a teacher is simultaneously trying to give out good grades and achieve grades that are well spread out, then he will be using the grades he is awarding to two ends, and will have to balance the two objectives in making decisions about the grades he awards.

In this paper a model of a teacher responding to these incentives is presented and analyzed. Students are constrained in that they have one unit of time to split between working on an assignment and working on an exam, and they maximize the utility from their grade. They choose how to split their time between the two tasks. The grade on the assignment and the exam are either success or fail. Devoting more time to the assignment or the exam or having been endowed with more intelligence will increase the probability of success, but not guarantee a higher grade, thus capturing the chance aspect of university grading mechanisms. The teacher is constrained by the behavior of the students and maximizes a weighted sum of the mean of the grades and standard deviation of the grades. He chooses how to weight the two tasks in the basic model, and how to weight the two tasks and how difficult one task is relative to the other in the extensions.

The main result of this paper is that neither the teacher's incentive to achieve a high mean or the teacher's incentive to achieve a high standard deviation generally motivate him to have many grading mechanisms rather than only a few. In cases where the university has decided that teaching is more important, the students will achieve higher grades and a lower standard deviation. In cases where the university has decided that sorting is more important, the student will achieve lower grades and a higher standard deviation. Though the sacrifice of one for the other is not great. If teachers have control over the relative difficulty of different grading mechanisms, the grades may be even lower though they will also achieve greater sorting.

The remainder of the paper is organized as follows. The *Literature Survey* section gives background on the related literature on the topic and highlights the differences between it and my model. The *Model* section outlines the model. The *Model Analysis* section solves the model to a point, and does comparative statics. The *Numerical Example* section solves the model given certain functional forms and ranges of the parameters. The *Model Extensions* section solves the model where the relative difficult of the two tasks is a choice variable for the teacher to a point, does comparative statics, and provides a numerical example. The *Conclusion* section summarizes the finding and presents implications.

2. Literature Survey

2.1. Teacher Evaluations and Grades

Many papers have been written to investigate the relationship between teachers' grading standards and teacher evaluations. When teacher evaluations first became widely used, grades quickly increased, leading to the supposition that the teacher evaluations led to easier grading standards (Zangenehzadeh, 1988). As well, many of the papers that address the question found that there was a statistically significant positive relationship between higher grades and higher teacher evaluations, though it is usually found to be so small that it would not be a major concern of the teacher (Lichty, Vose, Peterson 1978). However, as Stratton, Myers, and King (1994) point out in response to the finding on the relationship between grades and teacher evaluation, using typical regressional analysis will only identify correlation, and it will be impossible to identify whether using teacher evaluations led to better teaching, and thus better student learning and higher grades, or whether teachers just lowered their standards to make getting higher grades easier to help their teaching evaluations. Litchy, Vose, and Peterson make the same point about how finding a relationship cannot settle which of the forces led to the correlation and proposed that a theoretical justification for the relationship between inflating grades and higher teacher evaluations was necessary in the discussion, though they were not the first to develop a theoretical model. Their model is an extension of The McKenzie Model, which they think is a good theoretical basis for explaining the impact of grade inflation.

The McKenzie model takes the student as a utility maximizer, who has a time constraint and whose utility is a function of the grades he achieves in his various courses and the leisure activities he spends his time on. The student can commit time to courses, which will increase his academic achievement, and commit time to leisure, which will increase his utility from leisure. The portion of time committed to each activity that the student chooses will be where the marginal rate of substitution (MRS) of leisure for grades equals the relative time cost, or, on a graph, where the indifference curve is tangent to the budget line. The theoretical explanation of how grade inflation can improve teacher evaluations is that if academic achievement is made easier, the time cost of academic achievement is lower. This is captured on the graph as a pivot of the budget line outwards. There are two effects of academic achievement becoming cheaper: the substitution effect and the income effect. The substitution effect is the effect of the relative prices changing, and is an influence for the student to substitute away from the more expensive good to the cheaper good. In this case, the substitution effect is an influence for the student to substitute away from leisure and towards academic achievement. The income effect is the effect of the purchasing power of the individual changing. With one good cheaper the student can buy more in total than he could before, and its influence on consumption of the good whose price changed is determined by the MRS. If the MRS is such that the income effect leads the individual to consume less of the good, then the good is an inferior good. In the McKenzie model, the case analyzed is where academic achievement is an inferior good and thus the substitution effect and the income effect work in opposite directions. If the substitution effect is not dominated by the income effect, the student will commit more time to school, and thus both get a better grade, partially because grading standards are lower and partially because they are committing more time to school, and learn more, because they are committing more time to school. Their satisfaction is greater in both terms of grades and in terms of learning. They will have greater satisfaction from the course, and reward the teacher with higher teacher evaluations. If the substitution effect is dominated by the income effect, then academic achievement is a Giffen good. A Giffen good is a good that, when it becomes cheaper (more expensive) the consumer buys less of it (more of it). In this case, the student commits less time to school than before. Even so, the student may still reward the teacher with higher teacher evaluations because they are getting more satisfaction from the combination of leisure and schooling.

They go on to say that if the university as a whole inflates its grades, either to elicit more effort from its students or to improve the students' satisfaction from school, it risks grade inflation progressing to the point where academic achievement becomes a Giffen good. If that occurs then the reputation of the school will be hurt, because the graduates are learning less than before but the grades will not reflect that. Employers will adjust their interpretation of the grads as a signal of student ability, and students will adjust their valuation of the signal, making academic achievement less valuable at all levels. This would lead to even less effort in school. It would not be in the school's interest to allow grade inflation to progress so far that academic achievement would become a Giffen good.

The McKenzie model is not the only model of its kind. Many papers investigate questions related to the students' problem usually involving the tradeoff between school and leisure. Papers by Becker (1982), Wetzel (1977), Kelley (1972), Kelley (1975), Neslon and Lynch (1984), and Staaf (1972) are some examples. The policy implication usually under investigation in these papers is how a change in the technological level of teaching effectiveness would impact student achievement and student learning. An assumption they all make is that time devoted into school improves achievement. This assumption is supported in a recent paper by Stinebrickner and Stinebrickner (2008). They overcome the usually problem of endogenity in the question of the relationship between studying and academic achievement: people who study more are generally more intelligent, and that is why studying more is correlated with higher grades. They use an instrumental variable (IV) for quantity of time studying. The IV chosen was whether or not the student's randomly assigned roommate brought a video game console with him, and thus the student's hours of studying would be negatively effected but it would not be correlated with any other characteristics of the student because the assignment of roommates was random. They find that hours of studying does have a statistically significant impact on academic achievement. The impact of studying an extra hour a day has the same impact on first semester grades as an increase of 1.40 standard deviations on the American College Test, which is the entrance exam for people entering college in the USA.

Though they find time studying does improve academic achievement, the link between teacher evaluations and grades likely hinges on teachers being able to elicit more effort. There is a paper by Wetzel (1977) that says that using a sample from Virgina CommenWealth University, he found evidence contradictory to the belief that teachers could affect the amount of effort a student devoted to school. He used the student's SAT scores as a proxy for effort in the course, using the rationale that people with higher SAT scores would learn at a quicker rate and therefore their effective effort, which would be a combination of effort and learning rate, would be higher. He found that nothing besides whether the student currently had a job was significant in determining the students effort, with special attention paid to the fact that both pre-course attitude towards the subject matter and post-course attitude towards the subject matter did not have statistically significant relationship with effort. This seems to indicate that the students' effort in a course cannot be affected by the teacher, whether in the form of better teaching techniques or technologies or awarding higher grades. This helps to explain the result that there seem to be not a strong relationship between grades and teacher evaluation because the teacher cannot impact the students' behaviour. This lessens the ability of the teacher to improve the learning or the relative achievement of the student, thereby limiting his ability to improve his satisfaction of the course.

This result supports the choice in my model that the students have a fixed amount of time to commit to school, and that the teacher must optimize within this constraint of the student. However, this result also leaves the answer to the question of why teachers would want to give high grades being that they perceive a relationship between high grades and good teacher evaluations despite it being only a weak relationship. Regardless of the actual relationship, if a teacher believes the relationship is generally strong, then he will be motivated to give out higher grades in the hopes of receiving higher teacher evaluations. The information the individual teacher generates is not enough to determine for himself whether the relationship is true, because he could always assume that his teacher evaluations would have been lower if the grades he gave out were lower.

2.2. Screening

There are two theories why people who attain higher levels of education get higher wages. The first is the human capital theory-- that people gain productive skills at university. The second is screening theory-- that employers use the information that the candidate has attained a higher level of education as a signal that he also has other desirable characteristics. Screening theory does not ignore the possibility of human capital accumulation. That's why Weiss (1995) states that screening theory subsumes human capital theory-- because it is just an addition to human capital theory. It says that whatever is true about human capital theory, it is also true that there is extra information available to the employer, above the fact that the candidate may have learned at school, with regard to the candidates' productivity. Screening theory is usually based on attaining higher levels of education rather than academic achievement in the level of education attained. However, this does not mean that academic achievement is not also a screening mechanism that is used by employers in determining productivity. In fact, the very presence of schools giving out grades would indicate that they are attempting to provide a signal of the ability of the graduates. Wiess (1983) made a model where both the level of education attained and the grade attained were used as signals, in an attempt to capture a more realistic picture of reality, which includes grades. The model setup is that each person chooses an amount of education to attain, and at the end of his education, he takes a test that is either passed of failed. People of higher ability have a higher probability of success on the test.

Generally, the theoretical explanation why other productive characteristic are correlated with higher levels of education is that people with these characteristics can complete the school at a lower non-monetary cost, i.e. smarter people can complete school with less effort. However, in this model the test serves as a mechanism to encourage only the high ability people to choose higher levels of education. Employers have a strategy that maps a level of education and grade on the test into a wage offer. The strategy awards success on the test at higher levels of education greater than success on the test at lower levels of education. If the student has chosen a high level of education, then it is a signal of him being high ability, because the student believes that he can succeed on the test. If the student succeeds on the test it is a signal that there is a higher probability of the student being high ability. The employers' expectations of individuals' productivity based on the level of education they chose and their grade on the test is self-fulfilling in the sense that the individuals, knowing the expectations of the employers, choose levels of education that fulfill the employers expectations. This paper supports the idea in my model that grades can involve chance and still, though only when interacting with other signals be an accurate signal. In my model, the frame is only for one course and the grades do not interact with another signal. The grade is a signal that there is a higher probability of the student being of high ability, but it is not completely accurate. If the frame was widened to include infinite classes or levels of education, the same model of one course could have grades being a useful signal.

Since grades are used as a screening mechanism, and screening performs important societal benefits, schools will have the incentive to perform the function of awarding grades properly. Screening is not just a tool used by people of higher ability to separate themselves for their own benefit of achieving a higher wage. It has two societal benefits according to Stiglitz (1975). The first is that of providing the correct wage to people to induce the efficient amount of labour. It is a distortionary "information wage tax" if a person is given an improper wage based on improper screening. People of higher ability should be given a higher wage because they are more productive, and then will be induced to devote more of their time to labour. The second societal benefit is the purpose of matching. It is better to have a high ability person doing difficult and important work and a low ability person doing easy and unimportant work rather than the other way around. It is also better to have a good plumber being a plumber and a good stockbroker being a stockbroker rather than the other way around. Schools allow people to identify both which jobs they are especially good at and how good they are at these jobs. It is a societal function of the schools, and they have incentive to properly perform this function so that students will trust that their work in school will be properly rewarded and employers will trust that the graduates are properly sorted. Otherwise, the school will fail. In my model the university rewards the teachers partially based on the standard deviation of the grades in their courses in an effort to achieve good sorting.

William Chan , Li, Hao , and Wing Suen address the question of what grades a university will give to its students when it is free to give whatever grades it chooses. Since, grades are used as signal of graduates' productivity, universities generally care about providing accurate signals of their students' ability, but not completely. The significance of the grades it gives its students comes from the relationship between the grades it gives and the wages its graduates are offered. Employers offer wages to the graduates based on the expected productivity of the graduate given the grade he received, accounting for the possibility that the university inflated the grades. There are two states of the world, which differ in the proportion of good students attending the university. In the good state, the university has a higher portion of good students and a lower proportion of mediocre students, than it does in the bad state. The university can either give out a high proportion of As, equal to the high proportion of good students there are in the good state, or a low proportion of As, equal to the low proportion of good student there are in the bad state. With two possible proportions of high ability students (H and L), and two possible proportions of As given out (h and l), there are four possible cases: [H,h], [H,l], [L,h], and [L,l]. In the first and the last of the cases, the university accurately reports the quality of its students. It gives out a proportion of As that matches the proportion of good students. In case two, it under-represents the ability of its students, which it would never be in its best interest to do. In case four, it over-represents the ability of its students, which could possibly improve the wages offered to its graduates.

The university cares about a weighted sum of the success of its good graduates and its mediocre graduates, with a disproportionally high weight on the good graduates. The rationale for this increased weight on the higher ability students is that the university's benefits from success exhibit increasing returns to scale. The idea is that it is better in terms of reputation and financial contribution to the school to have one Nobel Prize winner or one CEO, rather than an equal amount of fame or wealth spread among many graduates. The optimal strategy for the university, given certain conditions, in the case with one university or two or many, is to randomize and inflate grades with a positive probability. This improves the welfare of the mediocre students, who will sometimes get As when they deserve Bs, at the cost of hurting the good students who now get offered a wage that accounts for the probability that they are a mediocre student, but even with the increased weight on the good students, it is the optimal strategy for the school. Even with an incentive to properly sort, that incentive can be dominated by others, and schools may purposely give inaccurate signals.

In my model the teacher is bound to give fair grades, and can only affect the grades by putting more weight on the easier or the more difficult task. The teacher does not care about which students are receiving what grades, but only cares about sorting to the extent that the standard deviation is high. It is as though the teacher care that there is the potential for the student to be sorted into different groups, but not how they are sorted. The difference is that in William Chan , Li, Hao , and Wing Suen higher ability students always get better or equal grades in school, but in my model, since chance is involved, that is not always the case. The teacher cannot enforce that the higher ability student get better grades because they might not have earned them, but they had a better chance of earning them.

3. The Model

There is one teacher and a [0,1] continuum of students. Each student has an intelligence level of q. The probability density function of q is denoted by f(q), and is bound within $[q_L, g_H]$.

Each student will receive a grade for the course. The two tasks for the students are an assignment and an exam. The final grade a student receives is a weighted sum of the grade he receives on the assignment and the grade he receives on the exam, with the sum of the two weights equaling one. Thus, the grade a student receives is

$$M = \alpha M_a + (1-\alpha) M_e$$

where M_a is the student's grade on the assignment and M_e is the student's grade on the exam. The grade a student receives on a task is

$$M_i = \begin{cases} 1 \text{ with probability } qP_i(l_i) \\ 0 \text{ with probability } 1\text{-}qP_i(l_i) \end{cases} \quad i = a,e$$

where P_i is function that transforms l_i to a value within [0,1], and l_i is the amount of time the student committed to working on task i. The following are characteristics of the P functions: 1. $dP_i/dl_i > 0$, 2. $d^2P_i/dl_i^2 < 0$, 3. dP_a/dl_a evaluated at $l_a=0$ is infinity, and dP_e/dl_e evaluated at $l_e=0$ is infinity, and 4. $P_a(l) > P_e(l)$ for all values of l within [0,1]. The reasons for those characteristics are as follows: 1. + 2. probability of success on a task should increase with time spent on the task but at a decreasing rate, 3. the probability of success on a task if no time is spent on it should be zero, but it should be positive for positive amounts of time, and 4. the two tasks should be different is some respect, specifically, one should be more difficult to succeed at.

The utility function of each student is $u(M)$.The following are characteristics of the u function: $du/dg > 0$, $d^2u/dg^2 < 0$. Each student has one unit of time available to divide between working on assignments and working on exams. Since $dP_i/dl_i > 0$, $dM/dP_i > 0$, and $du/dM > 0$, the student will commit all of his one unit of time to the two tasks. Thus, the student's one choice variable is l_a, with the binding constraint requiring that $l_e = 1 - l_a$.

4. Model Analysis

Each student chooses a value of l_a to maximize his expected utility:

$$\begin{aligned} U &= E(u(M)) \\ &= u(1)qP_aqP_e + u(1-\alpha)(1-qP_a)qP_e + u(\alpha)qP_a(1-qP_e) + u(0)(1-qP_a)(1-qP_e) \end{aligned}$$

The utility function of the teacher is a weighted sum of the mean and the standard deviation of the grades, with the sum of the two weights equaling one. Thus, the utility function of the teacher is

$$
\begin{aligned}
T &= G[Mean] + (1 - G)[Std.Dev] \\
&= G\int_{q_L}^{q_H} E(M)f(q)dq + (1 - G)[(\int_{q_L}^{q_H} (1 - qPa)(1 - qPe)f(q)dq)(0 - Mean)^2 \\
&\quad + \int_{q_L}^{q_H} qP_a(1 - qP_e)f(q)dq])(\alpha - Mean)^2 \\
&\quad + (\int_{q_L}^{q_H} (1 - qP_a)qP_e f(q)dq)(1 - \alpha - Mean)^2 \\
&\quad + (\int_{q_L}^{q_H} qP_a qP_e f(q)dq)(1 - Mean)^2]
\end{aligned}
$$

The teacher chooses a value of alpha to maximize his utility. The teacher is constrained by the behavior of the students, in the sense the he cannot choose each student's l_a, but he can affect the students' choice of l_a with his choice of alpha. The teacher has perfect information, and, in particular, has information on how the students' choice of la is affected by, or is a function of, alpha. This is not unrealistic because teachers have likely learned general student behaviour from his past teaching experiences.

Recall, that

$$U = u(1)qP_a qP_e + u(1-\alpha)(1-qP_a)qP_e + u(\alpha)qP_a(1-qP_e) + u(0)(1-qP_a)(1-qP_e)$$

Setting the derivative of U with respect to l_a equal to zero yields the first order condition

$$[q^2 P_a' P_e + q^2 P_a P_e'][u(1) + u(0) - u(\alpha) - u(1-\alpha)] + qP_a'[u(\alpha) - u(0)] + qP_e'[u(1-\alpha) - u(0)] = 0$$

This equation implicitly defines the students' optimal choice of la denoted by la* as a function of alpha.

Notice that l_a^* is a corner solution for the values α=0 and α=1. Substituting α=1 into the FOC yields

$$qP_a'[u(1) - u(0)] = 0$$

and since $P_a' > 0$ and $P_a'' < 0$ (ie. P_a' is decreasing in l_a), the l_a that most satisfies the above equation is $l_a = \infty$, and thus the solution to the constrained maximization problem would be $l_a^* = 1$.Substituting **α**=0 into the FOC yields

$$qP_e'[u(1) - u(0)] = 0$$

and since $P_e' < 0$ and $P_e'' < 0$ (ie. P_e' is negative and is decreasing in l_a), the l_a that most satisfies the above equation is $l_a = -\infty$, and thus the solution to the constrained maximization problem would be $l_a^* = 0$.

It is also true that the solution to the FOC is never a corner solution for any values of alpha within (0,1). This is because both P_a' evaluated at l_a=0 is infinity, and P_e' evaluated at l_a=1 is infinity, and the student's objective function is increasing in both P_a and P_e.

Recall that

$$\begin{aligned} T &= G[Mean] + (1-G)[Std.Dev] \\ &= G\int_{q_L}^{q_H} E(M)f(q)dq + (1-G)[(\int_{q_L}^{q_H}(1-qPa)(1-qPe)f(q)dq)(0-Mean)^2 \\ &\quad + \int_{q_L}^{q_H} qP_a(1-qP_e)f(q)dq])(\alpha - Mean)^2 \\ &\quad + (\int_{q_L}^{q_H}(1-qP_a)qP_e f(q)dq)(1-\alpha - Mean)^2 \\ &\quad + (\int_{q_L}^{q_H} qP_a qP_e f(q)dq)(1-Mean)^2] \end{aligned}$$

where

$$\begin{aligned} Mean &= \int_{q_L}^{q_H} E(M)f(q)dq \\ &= \int_{q_L}^{q_H}[1(qP_a qP_e) + \alpha(qP_a(1-qP_e)) + (1-\alpha)((1-qP_a)qP_e) + 0(1-qP_a)(1-qP_e)]f(q) \\ &= \int_{q_L}^{q_H}[\alpha qP_a + (1-\alpha)qP_e]f(q)dq \end{aligned}$$

The mean is the expected grade of a student with a particular value of q integrated over all values of q. The expected grade of a particular student can be simplified to the weight on the exam multiplied by the student's probability of success on the exam, plus the weight on the assignment multiplied by the student's probability of success on the assignment.

Notice that the standard deviation term is composed of only four terms. That is because there are two tasks, an assignment and an exam, and two possible grades for each task, 1 or 0. There are only four possible grades that a student can receive for the course: 0, α, 1-α, 1. They are the results of the four possible combinations of the random grade determination processes for the assignment and for the exam, which are, respectively, fail-fail, success-fail, fail-success, and success-success. The proportion of students that receive a certain grade is the probability that an individual student with a particular q will receive that grade, integrated over all values of q. Each term in the standard deviation is the proportion of students that will receive the particular grade multiplied by the square deviation of that grade from the mean.

T can be simplified to

$$\begin{aligned} T = \; & G[Mean] \\ & + (1-G)[(\int_{q_L}^{q_H} qP_a qP_e f(q)dq)[(1-Mean)^2 + (0-Mean)^2 - (\alpha - Mean)^2 - (1-\alpha - Mean)^2] \\ & + \int_{q_L}^{q_H} f(q)dq(0-Mean)^2 \\ & + \int_{q_L}^{q_H} qP_a f(q)dq[(\alpha - Mean)^2 - (0-Mean)^2 \\ & + \int_{q_L}^{q_H} qP_e f(q)dq[(1-\alpha - Mean)^2 - (0-Mean)^2] \end{aligned}$$

Setting the derivative of T with respect to alpha equal to zero yields the first order condition

$$
\begin{aligned}
&G(dMean/d\alpha) \\
&+ (1 - G)[\int_{q_L}^{q_H} (qP_a' l_a' qP_e + qP_a qP_e' l_a') f(q) dq [2\alpha (1 - \alpha)] \\
&+ \int_{q_L}^{q_H} (qP_a qP_e) f(q) dq [2(1 - 2\alpha)] \\
&+ \int_{q_L}^{q_H} f(q) dq [2Mean(dMean/d\alpha)] \\
&+ \int_{q_L}^{q_H} qP_a' l_a' f(q) dq [(\alpha - Mean)^2 - (0 - Mean)^2] \\
&+ \int_{q_L}^{q_H} qP_a f(q) dq [2\alpha - 2Mean - 2\alpha (dMean/d\alpha)] \\
&+ \int_{q_L}^{q_H} qP_e' l_a' f(q) dq [(1 - \alpha - Mean)^2 - (0 - Mean)^2] \\
&+ \int_{q_L}^{q_H} qP_e f(q) dq [2(Mean + \alpha Mean + \alpha - (dMean/d\alpha) - 1] = 0
\end{aligned}
$$

where

$$dMean/d\alpha = \alpha q P_a' l_a' + (1-\alpha) q P_e' l_a' + qP_a - qP_e$$

The FOC does not necessarily determine the value of alpha the teacher chooses, because there are bounds on the variable α and the maximum could either be an interior solution or a corner solution. The solution to the teacher's problem could either be at the solution to the FOC or at the corner values of α=0 or α=1.

The student's l_a' is a function of q and α. Therefore, comparative statics can be used to understand how the student's l_a' changes with q and with α. The implicit differentiation formula can be applied to the student's FOC, and states that

$$dF = (dF/dl_a)dl_a + (dF/d\alpha)d\alpha = 0$$

which implies

$$dl_a/d\alpha = -(dF/d\alpha)/(dF/dl_a)$$

where

$$-(dF/d\alpha) = u'(1-\alpha)[-q^2 P_a' P_e + (1 - qP_a) qP_e] - u'(\alpha)[qP_a'(1 - qP_e) - q^2 P_a P_e']$$

and

$$dF/dl_a = [q^2 P_a'' P_e + 2q^2 P_a' P_e' + q^2 P_a P_e''][u(1) + u(0) - u(\alpha) - u(1-\alpha)$$
$$+ qP_a''[u(1-\alpha) - u(0)] + qP_e''[u(\alpha) - u(0)]]$$

The expression $-(dF/d\alpha)$ is negative. The expression dF/dl_a cannot be signed. However, for the cases where the utility function is linear the expression simplifies to

$$qP_a''[u(1-\alpha)-u(0)]+qP_e''[u(\alpha)-u(0)]$$

which is negative. In this case, the entire expression is positive.

This is the intuitive result. It is intuitive to think that as the weight on the assignment increases so would the student's choice of time to commit to the assignment. For a given choice of l_a, an increase in the teacher's choice to alpha makes the probability of success on the assignment more valuable and the probability of success on the exam less valuable. Since, before the change in the teacher's choice of alpha, the marginal benefit of increasing l_a was equal to the marginal cost of decreasing $1-l_a$, the change in alpha would upset this relationship. This would give the student an incentive to increase the time he commits to the assignment at the cost of decreasing the time he commits to the exam.

Implicit differentiation can be applied to the FOC to find an expression for dl_a/dq, specifically

$$dl_a/dq = -(dF/dq)/(dF/dl_a)$$

where

$$-(dF/dq) = (-1)(2q[P_a' P_e + P_a P_e'][u(1) + u(0) - u(\alpha) - u(1-\alpha)] + P_a'[u(\alpha) - u(0)]$$
$$+ P_e'[u(1-\alpha) - u(0)])$$

and

$$dF/dl_a = q^2[P_a'' P_e + 2P_a' P_e' + P_a P_e''][u(1) + u(0) - u(\alpha) - u(1-\alpha)] + qP_a''[u(\alpha) - u(0)]$$
$$+ qP_e''[u(1-\alpha) - u(0)]$$

neither of which can be signed. However, for the cases where the utility function is linear the expressions simplify to

$$-(dF/dq) = (-1)[P_a'[u(\alpha) - u(0)] + P_e'[u(1-\alpha) - u(0)]]$$

and

$$dF/dl_a = qP_a''[u(\alpha) - u(0)] + qP_e''[u(1-\alpha) - u(0)]$$

Again, in this case, dF/dl_a is negative. $-(dF/dq)$ is positive if the following condition is satisfied

$$P_a'[u(\alpha) - u(0)] > P_e'[u(1-\alpha) - u(0)]]$$

The condition can be interpreted as the marginal benefit of increasing l_a is greater than the marginal cost of decreasing 1-l_a. Since, according to the student's FOC these two terms are set equal to each other, and the numerator would be zero.

5. Numerical Example

5.1. Choices and Their implications

In this numerical example the following functional forms were used:

$$u(M) = M$$

$$P_a = l_a^{1/2}$$

$$P_e = 0.95 l_e^{1/2} = 0.95(1 - l_a)^{1/2}$$

and the distribution of q was set to uniform over the range [0.5,1].

The purpose of choosing $u(M) = M$ was to make the problem simple enough that the computer program would be able to solve it. Recall that the student's FOC is

$$[q^2 P_a' P_e + q^2 P_a P_e'][u(1) + u(0) - u(\alpha) - u(1-\alpha)] + qP_a'[u(\alpha) - u(0)] + qP_e'[u(1-\alpha) - u(0)] = 0$$

Only linear utility functions will have the property that $[u(1) + u(0) - u(\alpha) - u(1-\alpha)] = 0$. If that is not the case then l_a^* is a function of q. If l_a^* is a function of q, the teacher's FOC is much more complicated to solve, and, it seems, outside the capacity of the computer program that was used in solving the model. Linear utility might be just as justifiable as a concave utility function though. With consumption concave utility functions are usually used, due to people’s characteristic of decreasing marginal utility. However, with grades, student my not have decreasing marginal utility. There are different echelons of grades that students can be in depending on their grades, meaning that different grades are valuable for different reasons. Just below the passing grade, the extra few grade points are valuable to the student because he will earn the degree, however the next points after that are not nearly as valuable. Just below 80, the next few grade points are valuable to the student because he will earn the dean’s honour list award or graduating with distinction, which might be helpful for getting a job, but the next few grade points after that are not nearly as valuable. The same story can be told for students who are continuing their schooling and a few more grade points will get them into a better school, or for students who are winning awards or making good impression on their teachers who might provide references in their future. Approaching the boundary to each of these echelons, the student's utility function is convex, and just past the boundary the student’s utility function is concave. A good approximation to an upwards-sloping wave function is a straight line. Notice the straight line in the graph below better captures the movement of the true utility function than any concave function could.

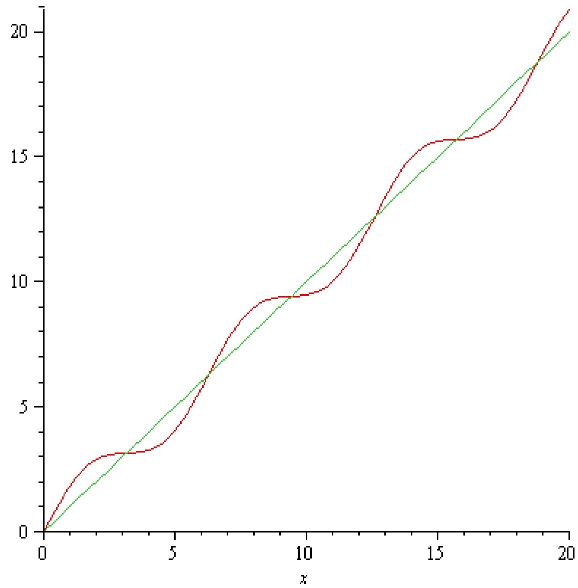

Figure 1: The above graph is sin(x) + x, as well as x

The functional forms $P_a = l_a^{1/theta}$ and $P_e = B(1-l_a)^{1/theta}$, where theta>1 and B<1, were chosen for two reasons. First, they satisfy all the desired characteristics outlined earlier which were $dP_i/dl_i > 0$, 2. $d^2P_i/dl_i^2 < 0$, 3. dP_a/dl_a evaluated at $l_a=0$ is infinity, and dP_e/dl_e evaluated at $l_e=0$ is infinity, and 4. $P_a(l) > P_e(l)$ for all values of l within [0,1]. Second, since they are meant to represent probabilities of success, they have the property that values of l_a within [0,1] will be transformed into values also within [0,1].

The choice of theta=2 was to keep the computations simple, and also was not so high that marginal effects of changing l_a would be difficult to detect. The choice of B=0.95 was so that weight on the assignment would improve class average and weight on the exam would improve standard deviation. If the value of B was low enough, then too high a proportion of the students would be failing with weight on the exam, which would decrease the standard deviation.

The choice of the distribution of q being uniform was for simplicity and to keep interpretation of the results easier. The choice for the range of q to be [0.5,1] was to maintain the teacher's tradeoff between a high mean and a high standard deviation. The teacher's desire for a high mean is meant to be a force to increase α teacher chooses, while the teacher's desire for a high standard deviation is meant to be a force to decrease the α the teacher chooses. However, the range of q can drastically effect the impact of the teacher's choice of α on the standard deviation. The possibilities for the range of q are [0,small], [0,middle], [0, big], [0,1], [small, small], [small, middle], [small,big], [small,1], [middle, middle], [middle, big], [middle,1], [big, big], and [big, 1].

The cases where the range of q extends from an interior value to another interior value are too variable to address, with the bottom limit and the top limit each affecting the optimal choice of α differently. However, cases where q extends from a boundary, 0 or 1, to any other value can be address. In fact, all the cases where the range of q extends from 0 to any other value can be addressed at the same time. The solution to the maximization of the standard deviation in any of these case to α=1, because the students with low values of q are causing there to be too many students achieving a grade of zero. The teacher responds to this by attempting to put weight on the grade of 1 by making it easier to achieve. For, example, if the range of q was [0,1] then if α=1 there would be an equal number of people with a grade of zero as with a grade of 1. All students would commit all their time to the assignment, giving all students a P_a=1. The only factor that would lead to student receiving different grades then would be their ability. A student with a q=1 would be guaranteed a grade of 1, while a student with a q=0 would be guaranteed a grade of 0. With all students having a P_a=1 and with the entire weight on assignments, the class mean would be

$$\int_0^1 q\,dq = 0.5$$

Since with full weight on the assignment, an equal number of students achieve a grade of zero as receive a grade of one. This would result in the highest possible standard deviation. If the teacher puts full weight on the exam, which is more difficult to achieve success in, there would be more people with a grade of 0 than with a grade of 1. Thus over this range of q, the highest standard deviation is achieved with α=1.

Skipping the cases where q extends from a small or middle value to 1 for a moment, because that is the case in the numerical example, if the range of q was [0.99,1], then when there is heavy weight on one task, whichever it is, and the students can specialize, too many will achieve a grade of 1. Though there will be some who achieve a grade of 0, the high percentage of students with a grade of 1 and the mean so close to 1 will lead to a low standard deviation. The standard deviation can be improved with an interior solution of α because though the points where the weight is placed are more tightly grouped, the weight will be spread more evenly across the points, leading to a higher standard deviation. In the graph below, the range of q was set to [0.99,1], and notice how interior values of alpha achieve greater standard deviation than corner values. Though this was not the case used in the numerical example, because in reality ability is more varied across student than this, the implications of it will be briefly discussed later.

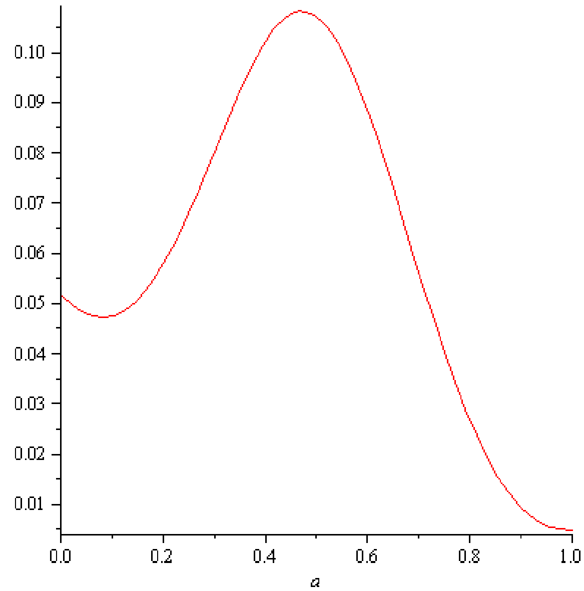

Figure 2: alpha against standard deviation with the range of q [0.99,1]

In this numerical example the range of q is [0.5,1]. In contrast to the example where the range of q was [0,1], this time if α=1 the mean would be

$$\int_{0.5}^{1} 2q\,dq = 0.75$$

which would mean there are three times as many student achieving a grade of 1 as are achieving a grade of 0. Thus the teacher, wanting to increase the standard deviation, has the incentive to make achieving a grade of 1 more difficult by shifting weight to the exam. In this case, the q extends low enough that there are enough students achieving a grade of zero that the maximum of the standard deviation is achieved at α=0.

The range of q[0,1] was rejected in favour of the range of q[0.5,1] to avoid the maximum for mean and standard deviation being achieved at the same value of α. Of the two situations realized by the two ranges of q, the latter more captures the problem faced by teachers, and more addresses the questions this paper is attempting to answer.

5.2. Results

Recall, the student's FOC

$$[q^2 P_a' P_e + q^2 P_a P_e'][u(1) + u(0) - u(\alpha) - u(1-\alpha)] + qP_a'[u(\alpha) - u(0)] + qP_e'[u(1-\alpha) - u(0)] = 0$$

Substituting the functional forms used in this example yields a closed formed solution for the student's optimal choice of l_a, specifically

$$l_a^* = \frac{\alpha^2}{(0.95(1-\alpha))^2 + \alpha^2}$$

The student's optimal l_a is a function only of α and not of q. The graph of l_a as function of α, increases slowly with α, then quickly, then slowly. That is to say that the marginal effect of an increase in alpha is greater for the middle values of α, and is smaller at lower and higher values of α. The story behind the graph is that the student chooses to specialize in one task or the other; when one task has much more weight on it he is slow to commit time to the other task, but, as the weight on the other task begins to overtake the weight on the first task, he quickly shifts time to that task. This is the result of the concavity of the P_a and P_e functions and the success-failure aspect to the grading system. The concavity of the probability of success functions P_a and P_e encourages diversification of time commitment rather than specialization. The graph on the left uses the functions $P_a = l_a^{1/2}$, $P_e = 0.95(1-l_a)^{1/2}$, and the graph on the right uses the functions $P_a = l_a^{1/5}$, $P_e = 0.95(1-l_a)^{1/5}$. Notice that the graph on the right increases more steadily with α. That is because those P_a and P_e functions are more concave, which rewards diversification of time commitment more highly.

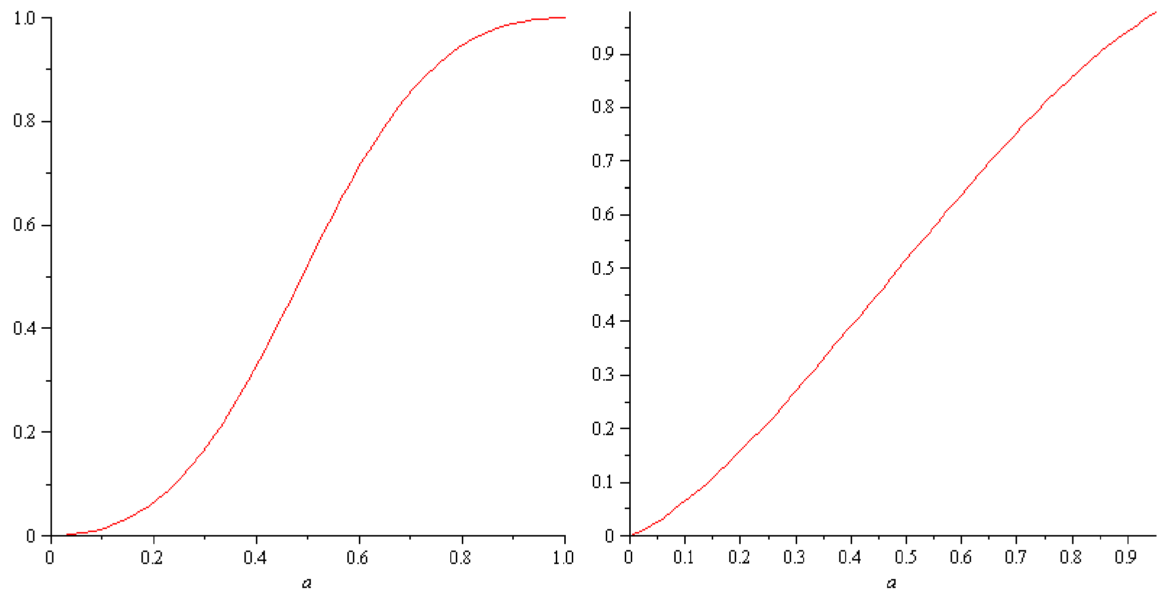

Figure 3: comparison of la* where probability of success functions have different levels of concavity

The success-failure aspect to the grading system encourages specialization of time commitment. For small weights on one of the two tasks, the student is unwilling to commit much time to that task because he must sacrifice the probability of success on the higher weighted task multiplied by the weight on that task. Though the probability of success is not affected much, the weight on the task is great enough that the student chooses to avoid the risk of failure on the higher weighted task. The two forces are at competition with each other in determining the shape of the curve, and though the success-failure aspect of the grading system cannot be altered for the sake of exhibition, when the concavity of the probability of success function is increased, the strength of that force becomes more powerful in comparison to the other and the effect was shown.

Neither the mean nor the standard deviation have a local maximum in interior values of α over the range of α [0,1]. With regard to the mean, there is a local maximum at $\alpha=0$ and at $\alpha=1$, with the $\alpha=1$ maximum achieving a greater value of the mean than the value $\alpha=0$

achieves. With regard to the standard deviation, the is a local maximum at α=0 and at α=1, with the α=0 maximum achieving a greater value of the standard deviation than the value α=1 achieves.

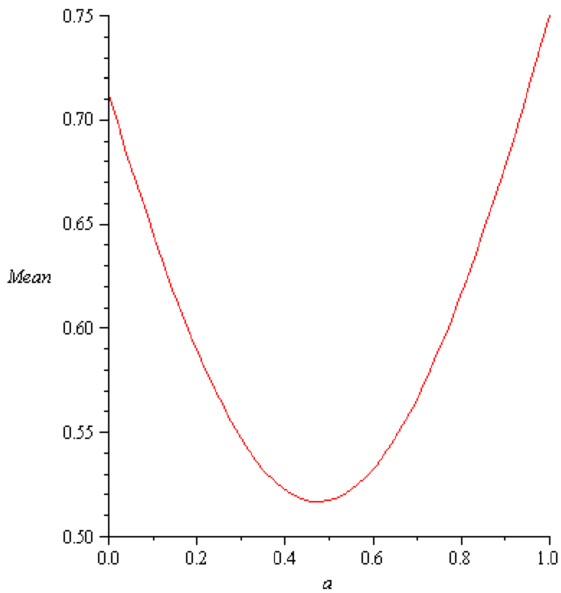

Figure 4: mean against alpha

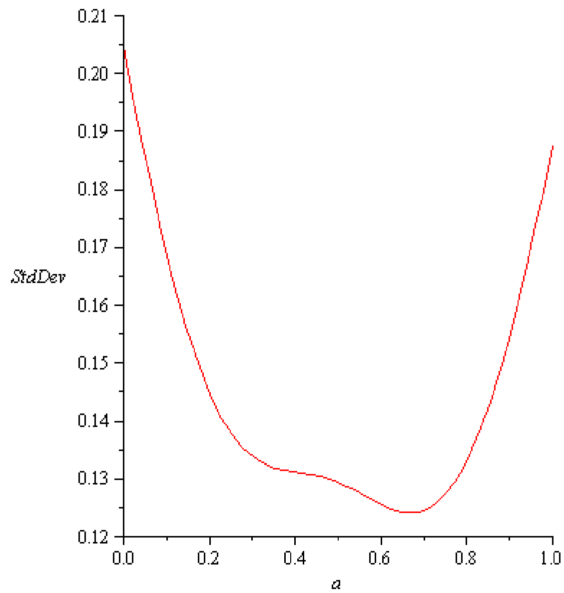

Figure 5: standard deviation against alpha

The intuition for the two shapes is as follows. The mean is improved when there is heavy weight on one task rather than the weight being spread between the two tasks. For example, if all the weight is on the assignment, ignoring the presence of q, a student's probability of success is 100%. If a tiny portion of the weight is shifted to the exam, then the probability that the student will succeed on the exam is much lower than the probability that the student will succeed on the assignment. The weight was shifted from where the student had a very high probability of success to where the student had a very low probability of success. The same logic applies to moving from a case where all the weight is on the exam to where a tiny portion of the weight is on the assignment. This shows that the teacher can improve the expected grade by allowing the student to specialized, that is, to commit all of his time to either one task or the other rather than splitting it between the two tasks. The remaining choice for the teacher is whether to allow the student to specialize on the assignment or the exam. Since the assignment is easier, in the sense that the same of amount of time committed to both the assignment and to the exam will yield a higher probability of success on the assignment, the local maximum at $\alpha=1$ is greater than the local maximum at $\alpha=0$.

The standard deviation is improved when there is heavy weight on one task rather than the weight being spread between two tasks as well. When the weight is entirely on one task, there are only two possible grades that a student can receive. That is to say that all students will receive either a grade of 0 or 1. When the weight is spread between the two tasks, there are four possible grades that a student can receive. That is to say that all the students will receive either a grade of 0, α, $1-\alpha$, or 1. For the teacher there are two effects of on the standard deviation of his choice of α. First, he is deciding at what points along the grade distribution within [0,1] there will be weight. That is to say, he is deciding what the possible grades are. Second, he is indirectly deciding what the weight at each of the point will be. That is to say, he is indirectly deciding what proportion of students will achieve each grade. The teacher's objective in these two decisions is to, as much as he can, simultaneously set the possible grades far away from the mean and to have the greatest weight on the grades that are the furthest from the mean. With this range of q, and the significance of it being discussed earlier, the maximum at $\alpha=0$ is greater than any interior value or the maximum at $\alpha=1$.

The maximum of the mean function is at $\alpha=1$. When there is full weight on the mean, ie when G=1, the solution to the teacher's problem is $\alpha=1$. The maximum of the standard deviation function is at α=0. When there is full weight on the standard deviation, ie when G=0, the solution to the teacher's problem is α=0. Since neither the mean nor the standard deviation has an interior maximum over the range of α [0,1], no linear combination of the two will have an interior maximum over the range. So, regardless of the choice of G, the solution to the teacher's problem is never an interior solution. That is to say, the only solutions are α=0 and α=1. These facts together imply that there must be a threshold value of G where, below that value, the solution to the teacher's problem is α=0, and above that value, the solution to the teacher's problem is α=1. In this numerical example, the threshold value is near G=0.3162. At that point the teacher jumps from an α=0 to an α=1.

The welfare of the students and teacher change with the parameter G. For the student, G does not directly affect his utility, but only indirectly because it affects the teacher's choice of α. However, since there are only two values of α that the teacher ever chooses, there are only two levels of welfare that the students ever achieve, and they are the welfare associated with an α=0 and a welfare associated with a α=1. Since the utility function is linear, the welfare of each student is the expected grade of the student. The expected grade of each student integrated over all students is the mean. The mean when α=1 is 0.75. The mean when α=0 is 0.7125 (=0.75*0.95). The student's welfare is greater if G is past the threshold value of G=0.3162 . That is to say, that if the teacher's incentives are weighted heavily enough on the mean, then the student's welfare will be greater.

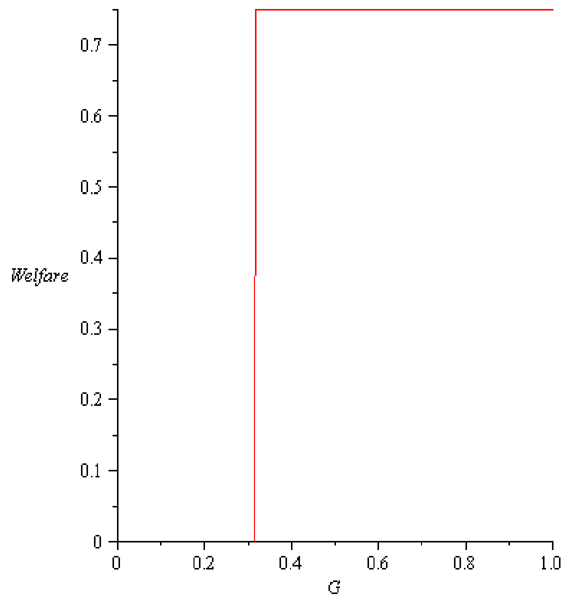

Figure 6: student welfare against G

The teacher's welfare changes with G as well. Tough, the teacher's utility changes directly with G. As the weight on mean versus standard deviation changes, the teacher utility increases as the mean is always greater than the standard deviation for both values of α. For values of G between 0 and 0.3162 the teacher's choice is α=0 which achieves a standard deviation of 0.2048 and a mean of 0.7125. As G increases over that range the teacher receives more of the benefit from the mean being greater than the standard deviation. For values of G between 0.3162 and 1, the teacher's choice of α=1 which achieves a standard deviation of 0.1875 and a mean of 0.75. As G increases over that ranges the teacher receives more the benefit from the mean being great than the standard deviation, and since the mean as a proportion of the standard deviation is greater for α=1 than for α=0, the teacher's welfare rises faster with G. The teacher's welfare then is the max of the two curves, one being the teacher's welfare when a=0 and the other being the teacher's welfare when a=1. Notice that they cross at G=0.3126 where α=1 overtakes α=0.

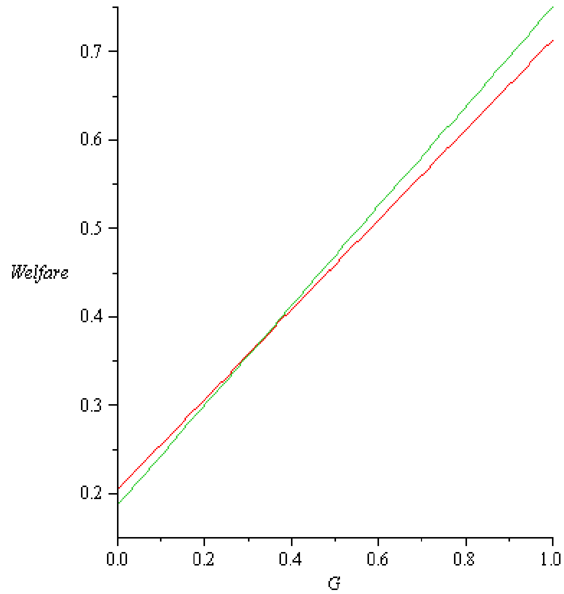

Figure7: teacher's welfare at alpha=0 and alpha=1 against G

In reality though, it would be strange if the teacher benefited more or less depending on whether the university was pursuing good teaching or good sorting. It would most likely be the case that, if the teacher was rewarded according to these two goals, it would be some more complex function of the two. The teacher could be indifferent between any of the possible weights on mean versus standard deviation, and still exhibit the same behavior with respect to choosing α.

Comparative statics in this numerical example are very simple. The derivative of l_a with respect to α has already been discussed. Since, the students' l_a only changes in response to α, and doesn't change except to jump from 0 to 1 around G=0.3162, l_a only changes with G to jump at the same value of G, from the l_a=0 associated with α=0 to the l_a=1 associated with α=1. The graph is similar to the graph of the welfare of students in that it takes one value up to G= 0.3162 where there is a break and then takes another value afterward.

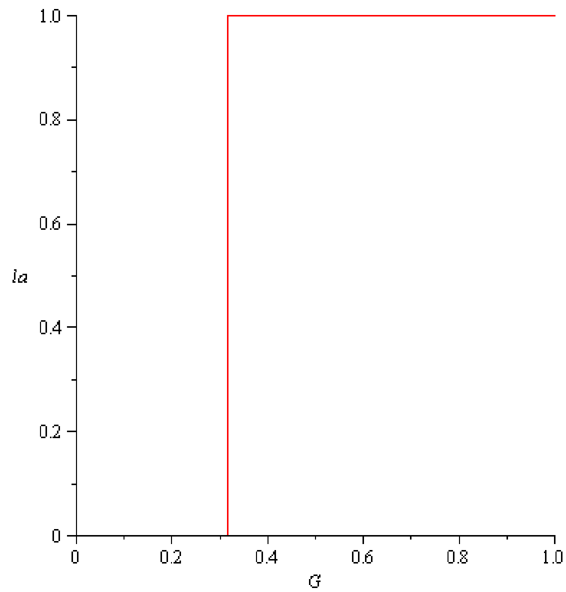

Figure8: student's la* against G

Returning to the case where the range of q is [0.99,1], we might expected there to be two jumps in the teacher's choice of α rather than just one. The initial maximum, for G=0, would be at an interior value of α. At G=1 the maximum would still be at α=1. So there will at least be one jump. However, since both the class average and the standard deviation have a local maximum at α=0, there may be a jump from an interior value to α=0, and then another jump from α=0 to α=1.

6. Model Extensions

6.1. Model Analysis

In the basic model the functions P_a and P_e were not given functional forms. Then, in the numerical example of the basic model, they were given function forms, and the coefficient used in the P_e function was constant. In this model the functional forms of P_a and

P_e are defined, just the same as they were in the numerical example, except that the coefficient used in P_e is variable and is the teacher's choice variable. In this model $P_a = l_a^{1/2}$ and $P_e = B(1 - l_a)^{1/2}$. B is a measure of how much more difficult the exam is relative to the assignment.

The students' objective function is

$$U = u(1)ql_a^{1/2}Bq(1 - l_a)^{1/2} + u(\alpha)ql_a^{1/2}(1 - Bq(1 - l_a)^{1/2}) + u(1 - \alpha)(1 - ql_a^{1/2})Bq(1 - l_a)^{1/2}$$
$$+ u(0)(1 - ql_a^{1/2})(1 - Bq(1 - l_a)^{1/2})$$

The first order condition is

$$q^2B(1-2l_a)[u(1)+u(0)-u(\alpha)-u(1-\alpha)]+q(1-l_a)^{1/2}[u(\alpha)-u(0)]+qBl_a^{1/2}[u(1-\alpha)-u(0)]=0$$

The implicit differentiation formula can be applied to find an expression for $dl_a / d\alpha$, specifically

$$dl_a / d\alpha = (-dF/d\alpha)/(dF/dl_a)$$

where

$$-dF/d\alpha = (1/2)[u'(1-\alpha)[-q^2l_a^{1/2}(1-l_a)^{1/2} - (1-ql_a^{1/2})(1-l_a)^{1/2}] - u'(\alpha)[ql_a^{1/2}(1-(1-l_a)^{1/2})$$
$$+ q^2l_a(1-l_a)^{1/2}]$$

and

$$dF/dl_a = Bq[u(\alpha) + u(1-\alpha) - u(1) - u(0)] - (1-l_a)^{3/2}[u(\alpha) - u(0)] - Bl_a^{3/2}[u(1-\alpha) - u(0)]$$

Just as it was without function forms specified, $-dF/d\alpha$ is negative, and dF/dl_a cannot be signed, but, if the utility function is linear, it will be negative, making the expression in total positive.

The implicit differentiation formula can be applied to find an expression for dl_a / dq, specifically

$$dl_a / dq = (-dF/dq)/(dF/dl_a)$$

where

$$(-dF/dq) = 2Bq(1-2l_a)[u(\alpha)+u(1-\alpha)-u(1)-u(0)] - (1-l_a)^{3/2}[u(\alpha)-u(0)]$$
$$- Bl_a^{3/2}[u(1-\alpha)-u(0)]$$

and

$$dF/dl_a = Bq[u(\alpha)+u(1-\alpha)-u(1)-u(0)] - (1-l_a)^{3/2}[u(\alpha)-u(0)] - Bl_a^{3/2}[u(1-\alpha)-u(0)]$$

Again, as without specified functional forms specified, neither can be signed, and in the linear utility case, -dF/dq=0, because the condition for it to be positive(negative) is that the FOC be positive(negative).

Since there is a new variable, B, the implicit differentiation formula can be applied to find an expression for $dl_a dB$, without functional forms first, specifically

$$dl_a / dB = (-dF/dB)/(dF/dl_a)$$

where, with Pj=Pe/B,

$$-dF/dB = [q^2 P_a^{'} P_j + q^2 P_a P_j^{'}][u(\alpha)+u(1-\alpha)-u(1)-u(0)] - qP_j^{'}[u(1-\alpha)-u(0)]$$
$$= q^2 P_a^{'} P_j [u(\alpha)+u(1-\alpha)-u(1)-u(0)] + q^2 P_a P_j^{'}[u(\alpha)-u(1)]$$
$$+ q[qP_a - 1]P_j^{'}[u(1-\alpha)-u(0)]$$

and

$$dF/dl_a = Bq[u(\alpha)+u(1-\alpha)-u(1)-u(0)] - (1-l_a)^{3/2}[u(\alpha)-u(0)] - Bl_a^{3/2}[u(1-\alpha)-u(0)]$$

$-dF/dB$ is non-negative, because, with a linear or concave utility function, all three terms in $-dF/dB$ are non-negative. With a linear utility dF/dl_a is negative, and the sign of dl_a / dB would be negative. This is intuitive because as the exam becomes easier to succeed in, that is, as the time committed to it receives a greater reward in terms of probability of success, the student has incentive to increase the amount of time he commits to the exam.

6.2. Numerical Example

The functional forms are the same as the ones used in the basic model, except rather than a coefficient in the expression for P_e, a variable is used, so the functions are

$$u(M) = M$$

$$P_a = l_a^{1/2}$$

$$P_e = B(1 - l_a)^{1/2}$$

and the distribution of q was set to uniform over the range [0.5,1].

Substituting the functional forms used in this example yields a closed formed solution for the student's optimal choice of l_a, specifically

$$l_a = \frac{\alpha^2}{(B(1-\alpha))^2 + \alpha^2}$$

To give the reader an understanding of how the student's la and how the teachers objective function are affected by α and B, without including 3D graphs, a few graphs have been included. For a particular value of α=0.75, the graph of l_a against B looks like this

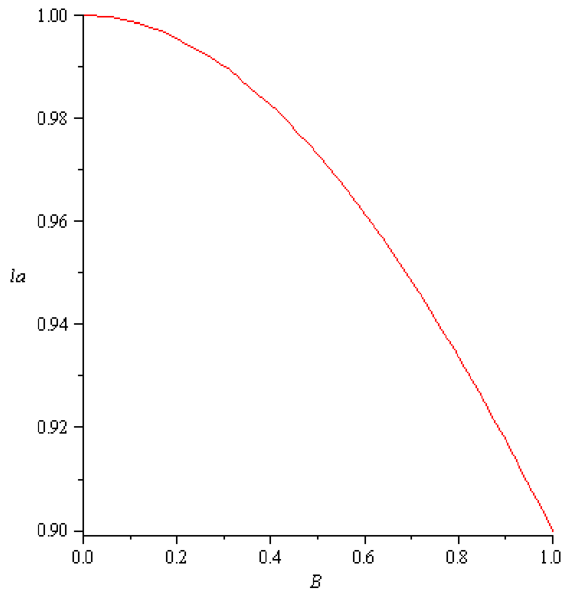

Figure9: students' la* against B

For a particular value of B=0.75, the graph of l_a against α looks like this

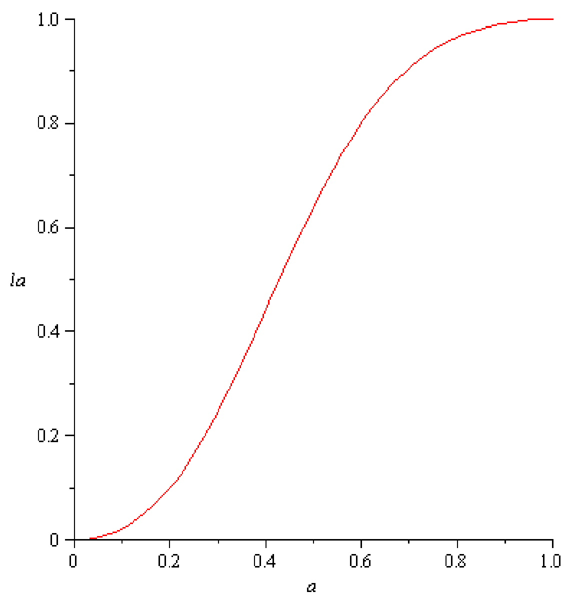

Figure10: students' la* against alpha

The graph of the teacher's objective function against B, for a particular value of G=0.2, and particular α=0, looks like this

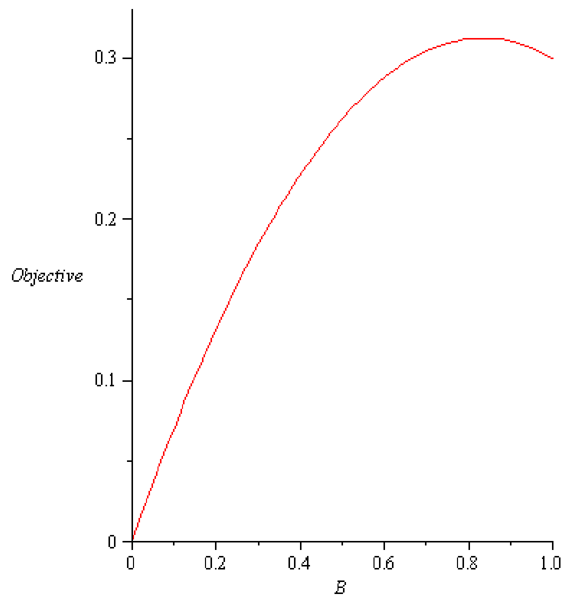

Figure11: teacher's objective against B

The graph of the teacher's objective function against α, for a particular value of G=0.2, and particular B=0.85, looks like this

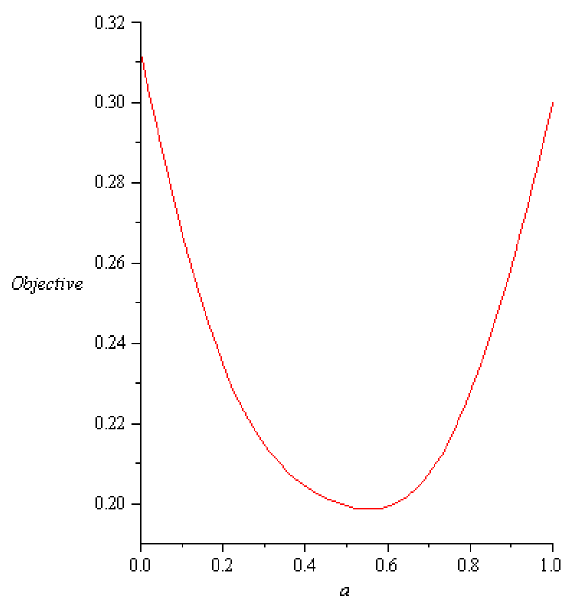

Figure12: teacher's objective against alpha

The teacher's objective function is still such that the teacher would never choose an interior value of α for all values of B, but it is concave in B for all values of α. Thus, the result from the base model, where the teacher only ever chooses an α of 0 to 1 remains true in this extended model. However, the teacher does, for values of G within a certain range, choose interior values of B.

There is an interaction between these two choice variables that is of importance, and should be kept in mind. When α=1, the teacher choice of B is irrelevant, because there is no weight on the exam so it doesn't matter how difficult it is. When B=1, whether the teacher chooses α=0 or α=1 is irrelevant because the exam and the assignment are equally difficult. Since, when B=1, the teacher's objective function is symmetric in alpha, only the magnitude of the two weights matters and not which task has which magnitude.

The expected grade is convex in α but always decreasing in B. The expected grade can never increase with B because an increase in B only makes success in the exam more difficult to achieve, and thus less people achieve it.

The optimal choices of α and B change with G. Since the maximum of standard deviation is achieved with α=0 and an interior value of B, when G=0, this is the solution. As G increases there is more weight put on the expected grade, which is decreasing in B, thus the solution shifts to a higher value of B, remaining with an α=0. B increases with G until the point G=0.334, when the maximum Beta=1 is reached. At this point, both the exam and the assignment are equally difficult to achieve success in and thus the choice of α=0, 1 is equivalent. After this point, the solution remains B=1 α=0 or 1, because this is the maximum for all values of G after this point.

The optimal choice of B is a function of G, for all G within [0,1], and is

$$B = \frac{(-2/3)}{(-1+G)}$$

when that expression is less than the $B \le 1$ boundary, and

$$B = 1$$

when that expression is greater than or equal to 1.

The graph of B against G looks like this

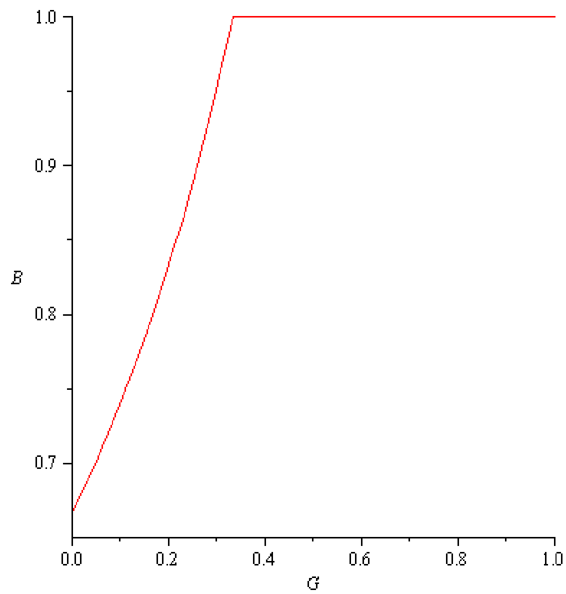

Figure13: teacher's B* against G

The graph of the mean against G looks like this

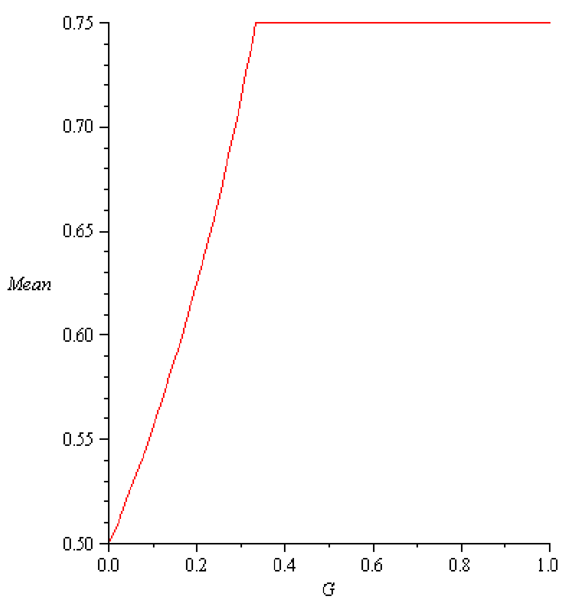

Figure14: mean against G in the extended model

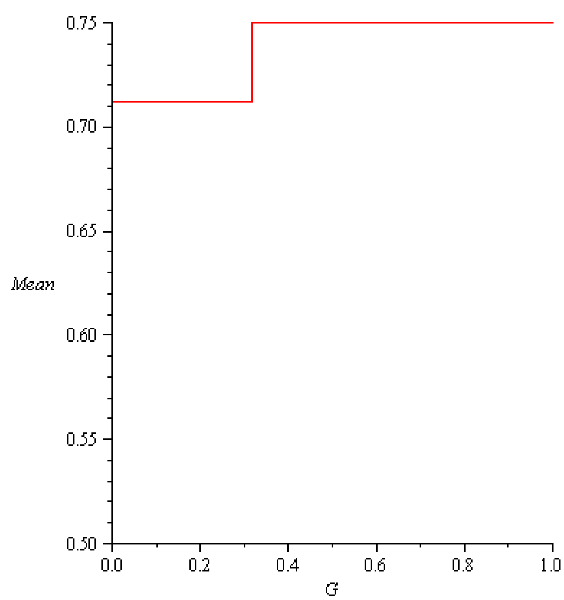

Figure15: mean against G in the basic model

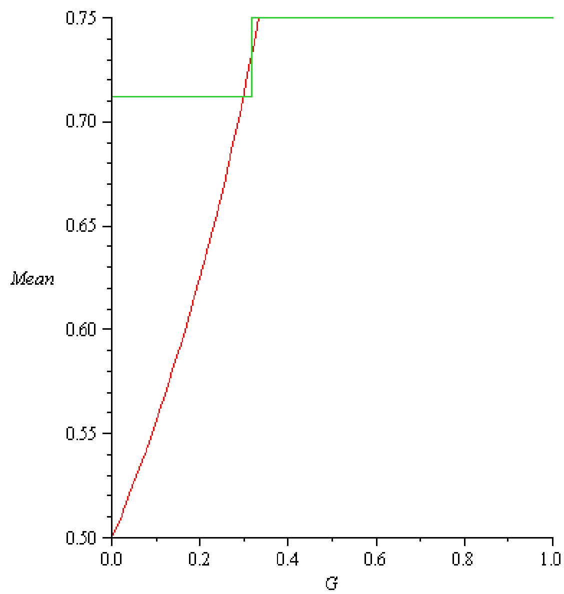

Figure16: mean against G for both models

For values of G below 0.2982, the extended model achieves a lower mean. At G=0.2982, the teacher's choice is B=0.95, which is the fixed values of B in the basic model, and both models achieve the same mean. For the very small range between G=0.2982 and G=0.3162 the extend model actually achieve a higher mean than the basic model. This is because the teacher would raise the B marginal as G changes if he could, but in the basic model he only has the choice between B=0.95, by putting all the weight on the exam, and B=1, by putting all the weight on the assignment. Though he would raise B marginally if he could, α =0 still achieves a greater value of the mean than α =1. It is at the midpoint between G=0.2982 and G=0.3333, roughly, that it α =1 achieves a higher value of the mean in the basic model. From G=0.3333 to G=1 they achieve the same mean, though this is only because B is bound within [0,1], and so it cannot continue its upward trend. In reality, this upper bound could be interpreted as a minimum difficulty level of the course demanded by the university.

The relationship between the two means is based largely on the fixed B that was chosen in the basic model. For example, if the B in the basic model was 0.80, then the extended model would have a lower mean for value of G below G=0.1667. Then over the range of G=0.1667 to G=0.2500 the basic model would have a higher mean, because the B would rise in the extended model, but in the basic model the G would not have reached the value where α jumps from α =0 to α=1. Over the range of G=0.25 to G=0.3334 the extended model would again have a lower mean. After G=0.3334 the two means will be the same. The interpretation of the fixed value of B in the basic model is how much less likely it is the students will succeed on exams compared to assignments when it is out of the teacher's control. Since, it likely is in the teacher's control, it is impossible to say what a likely value of the fixed value of B in the basic model should be. Thus, comparisons between the magnitudes of means and standard deviations achieved in two models are more or less meaningless.

The graph of the standard deviation against G looks like this

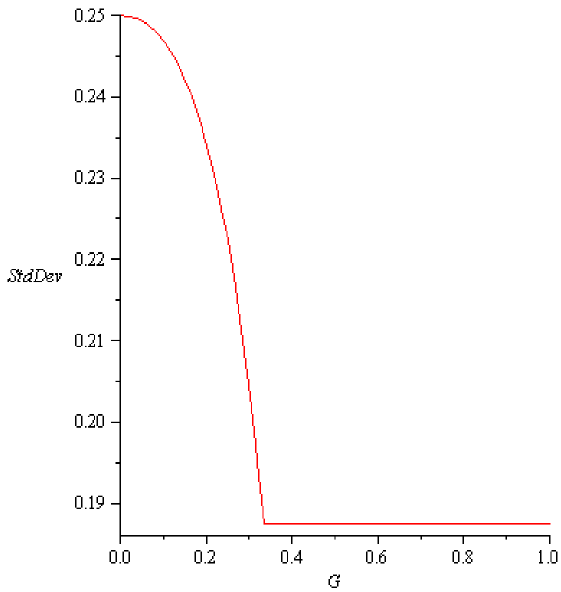

Figure17: standard deviation against G for the extended model

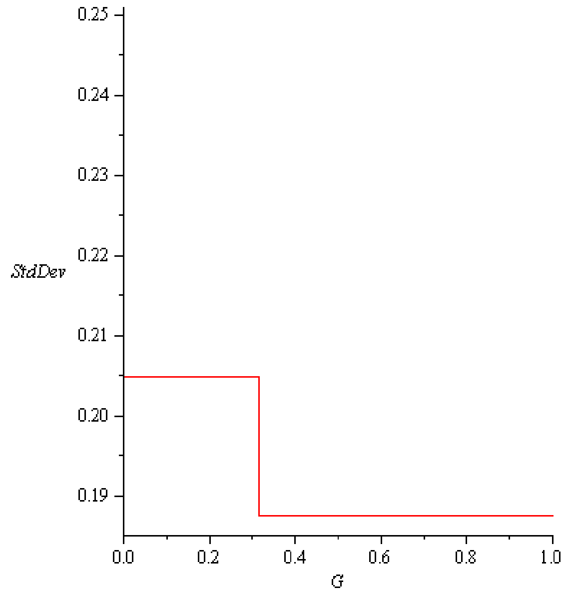

Figure18: standard deviation against G for the basic model

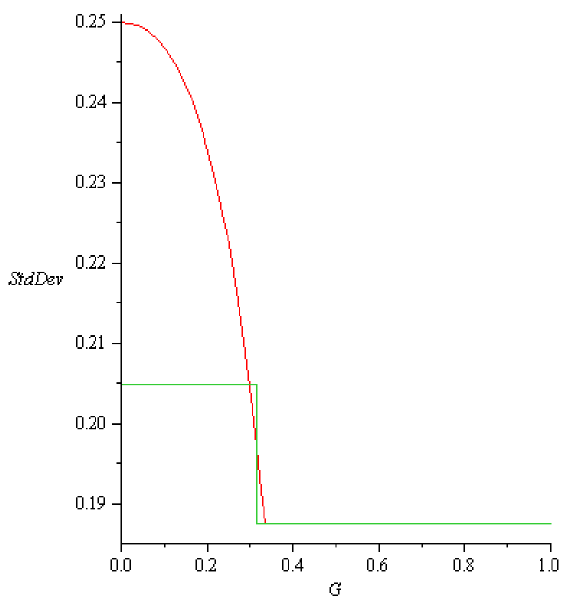

Figure19: standard deviation against G for both models

Again, it is at G=0.2982 that the extended model and the basic model achieve the same value of standard deviation. For values of G below that the standard deviation is much higher in the extended model, and there is only a small range over which the standard deviation is higher in the basic model.

The graph of student welfare against G is increasing for the range of G where Beta is identical to the graph of the mean achieved because the students utility function is linear. Comparisons of students' utility in the two models suffer from the same problem as did comparisons of the mean.

Notice that the teacher achieves a higher welfare by choosing both B and α rather than just α where the B is fixed at 0.95. In both cases for all G>0.334, the teacher is putting a full weight on the assignment, or equivalently, on an exam that is set to be equally as difficult as an assignment. However, before that range of G, the teacher achieves higher levels of welfare by altering the B he chooses.

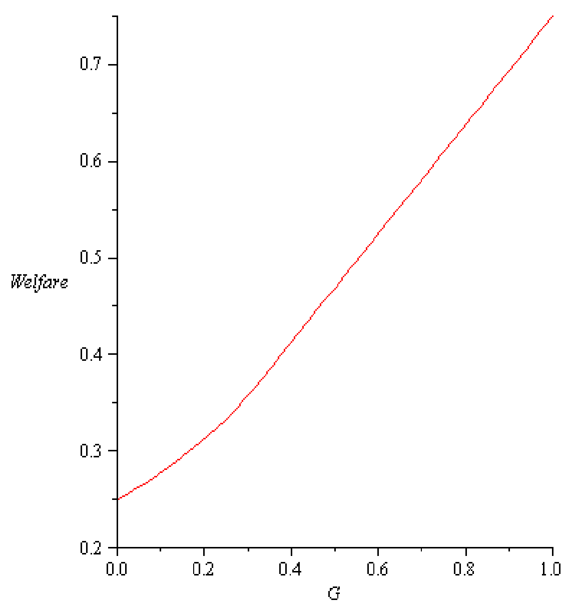

Figure20: teacher's welfare against G for the extended model

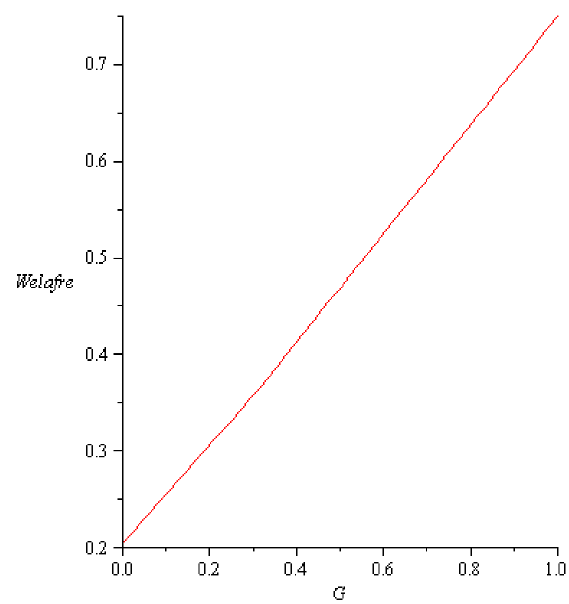

Figure21: teacher's welfare against G for the basic model

Notice of the graph below, how the line representing the teacher's welfare when he chooses B is always greater than or equal to the line representing the teacher's welfare when

B is fixed at 0.95. This is because the teacher can always do as well, by choosing B=0.95, and would only deviate from this choice if it improved his welfare.

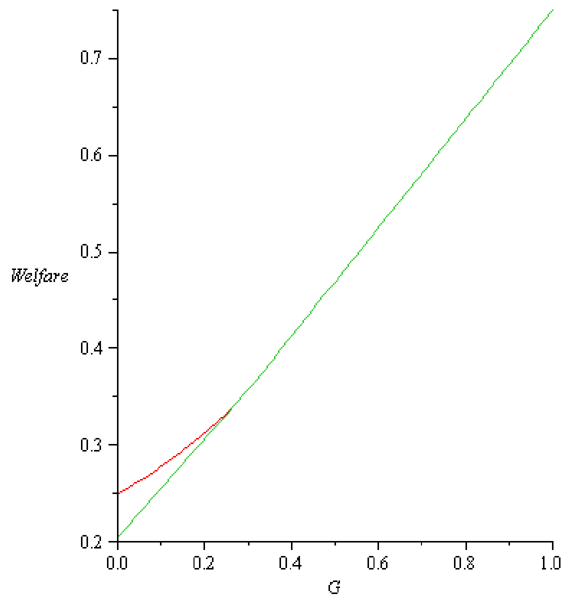

Figure22: teacher's welfare against G for both models

7. Conclusion

G is the parameter of focus in this paper. In the introduction, G was explained to be the relative importance in the university program of having good education versus having good sorting. The university then achieved the optimal balance of these two goals by rewarding its teacher with a weighted sum of teacher evaluations and the class standard deviation. The teacher's objective with respect to grades was to achieve a high mean to elicit good teacher evaluations and achieve a high standard deviation. It will be helpful to return to the main results of the paper and interpret a high G to be the undergraduate program case and a low G to be the PhD case.

In an undergraduate program you would expect there to be a higher weight on teaching evaluations rather than class grade standard deviation. This leads the teacher to put all the weight on the assignment. The students get better grades, because success in the course has been made easier. Even if the teacher has control over the relative difficulty between the assignment and the exam, since the weight on teacher evaluations is significantly high, the choice of the teacher can either not change or change to a grading standard that is identically difficult but with all the weight on an exam. The second tends to be the case in undergraduate programs, where there are very few grading mechanisms, and heavy weight on two or three exams over the course of a year. Though in reality there is more than one grading mechanism, the reality that there are few rather than many, matches the model predictions that there is one rather than two.

In a PhD program you would expect there to be a higher weight on the standard deviation rather than teacher evaluations. This leads the teacher to put all the weight on the exam. The students achieve a lower welfare because probability of success is lower. If the teacher has control over the relative difficulty between the assignment and the exam, then the exams will be even more difficult, but only assuming the difficulty level was relatively equal when it was outside the teacher's control, and a significantly low weight on teacher evaluations. This does not tend to be the case in PhD programs, where they have a few big papers to write rather than exams, but exams are just meant to be interpreted as the more difficult of the two tasks, and since big papers are often even more difficult than exams or assignments, the result can mesh well with reality after all. In PhD program, the students do have a few very difficult tasks.

The topic of this paper was an investigation into the determination of grades, which the sole function of are to be used as a signal of the ability of the student. Thus the implication of the paper must be about what this process reveals about the signal that grades are providing. The implication is that the grades awarded to the students are not fully in control of the university, the teachers, or the students, but it is the product of a process where the teachers are agents of the university and the students are the agent of the teachers. Since grades are meant to be a signal of the quality of a student, an accurate signal about a particular student would give the same impression to an employer whether the student was in a course that was difficult to succeed in or a course that was easy to succeed in. The information gained from this model should be used to signify that bad grades in undergraduate courses are much more likely to be indicative of low intelligence than bad grades in a PhD level course, because of the fact that PhD level courses are much more focused on sorting than undergraduate level courses, which, among other reasons, makes them more difficult.

Bibliography

Becker, W. E.,Jr. (1982). The educational process and student achievement given uncertainty in measurement. *American Economic Review, 72*(1), 229-236.

Chan, W., Li, H., & Suen, W. (2005). A signaling theory of grade inflation. http://www.economics.utoronto.ca/haoli/research/Schwarz.pdf

Kelley, A. C. (1972). Uses and abuses of course evaluations as measures of educational output. *Journal of Economic Education, 4*(1), 13-18.

Kelley, A. C. (1975). The student as a utility maximizer. *Journal of Economic Education, 6*(2), 82-92.

Lichty, R. W., Vose, D. A., & Peterson, J. M. (1978). The economic effects of grade inflation on instructor evaluation: An alternative interpretation. *Journal of Economic Education, 10*(1), 3-11.

Nelson, J. P., & Lynch, K. A. (1984). Grade inflation, real income, simultaneity, and teaching evaluations. *Journal of Economic Education, 15*(1), 21-37.

Staaf, R. J. (1972). Student performance and changes in learning technology in required courses. *Journal of Economic Education, 3*(2), 124-129.

Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *American Economic Review, 65*(3), 283-300.

Stinebrickner, R., & Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. http://economics.uwo.ca/faculty/stinebrickner/paper2.pdf.

Stratton, R. W., Myers, S. C., & King, R. H. (1994). Faculty behavior, grades, and student evaluations. *Journal of Economic Education, 25*(1), 5-15.

Weiss, A. (1983). A sorting-cum-learning model of education. *Journal of Political Economy, 91*(3), 420-442.

Weiss, A. (1995). Human capital vs. signaling explanations of wages. *Journal of Economic Perspectives, 9*(4), 133-154.

Wetzel, J. N. (1977). Measuring student scholastic effort: An economic theory of learning approach. *Journal of Economic Education, 9*(1), 34-41.

Zangenehzadeh, H. (1988). Grade inflation: A way out. *Journal of Economic Education, 19*(3), 217-226.