AFRICAN MICRO-ECONOMIC RESEARCH UNIT

Risk Aversion and the Gender Differential in Academic Performance

Ivor Abramowitz (LSE, AMERU)
Prof Neil Rankin (Wits, AMERU)
Mr Volker Schoër (Wits, AMERU)
Prof Corné van Walbeek (UCT)

September 2011

This paper estimates the gender differential in academic performance when assessment is through the use of multiple-choice format questions when there is a penalty for incorrect guesses. The paper uses a unique dataset, collected at the University of Witwatersrand, in Johannesburg, South Africa. Two measures of risk are included in the analysis in order to measure the effect of risk preferences on the gender differential in academic performance. The first is an individual-specific time-invariant measure of risk aversion and the second is a question-specific importance variable. This paper argues that pre-market selection is an important determinant of academic achievement. Women are, at an upper bound, 19 percent less likely to attempt a multiple choice question with such a format. Given that they do attempt it, they are 9 percent less likely than males to answer it correctly. There remains a performance differential for black students, independent of pre-market selection. The results suggest that, for females, learning experience as well as preferences for risk determine market participation and by implication, academic performance.

1. Introduction

Education economics literature that uses the education production function to analyse the academic returns to individual productive characteristics is usually embedded within Becker's (1964) model of human capital.

In an education setting, students are typically rewarded with the marginal product of their productive characteristics, both observable and unobservable (Becker, 1964). Productive human characteristics are modelled like inputs or investments in an economy and the returns are measured as the observed or estimated payoffs from those investments. Becker (1975) defines the returns, or outcome variable, as a function of observable individual productive characteristics such as age, gender and ethnicity. The outcome variable is usually academic achievement, pass rates, or graduation from school or university.

In most modern analyses in the human capital literature, econometric techniques are used to account for unobservable characteristics such as inherent ability and preferences. The literature mostly discusses ability and advanced econometric techniques are used, together with proxy variables or instrumental variables, to control for this. Research on another unobservable, discrimination, tends to ascribe the residual differential between male and female performance or minority versus majority ethnic groups as a lower bound of the effect of discrimination in the market (Oaxaca, 1973).

This paper argues that part of the differential between males and females can be accounted for by the inherent differences in the risk preferences held by students. Students who display higher degrees of risk aversion are less likely to answer multiple-choice questions. Since this testing format is commonly used in higher education, there reflects a gender differential in student performance.

Section 2 of this paper discusses both the theoretical and empirical literature on the topic. The South African context for the study and the data used in the analysis are described in sections 3 and 4 respectively. Section 5 provides the methodology behind the estimation strategy, while section 6 reports the results. The results are discussed in section 7, together with implication for further research. Section 8 concludes.

2. Literature

2.1. Human Capital Model

The economic importance of human capital cannot be understated. The study of education, healthcare, and social sector reforms are centred on the relationship between human capital and growth and development. The study of inequality and distribution of earnings and wealth in a country is directly founded in the inequality and distribution of education and medical care in that country (Becker, 1964). Becker's (1964) model of human capital, together with Mincer's (1970) human capital earnings function, provides a basis for the study on the relationship between education and the returns to education (Card, 1999).

In the model of human capital formation, Becker (1964) uses an investment approach. The amount as well as the type of capital invested is related to the rate of return expected from placing those resources in human capital. This expected return to the investment is based on a valuation that incorporates the economic, political and social opportunities as well as the social value of the productive output.

In modelling the returns to education, Card (1999) identifies that unobservable characteristics such as family background and ability are likely to bias the ordinary least square estimates of the returns. The size of this bias is a function of the heterogeneity of the sample and would be very difficult to measure. The direction of the bias, however, can be posited. Willis and Rosen (1979) argue that independent of the theoretical model that one evaluates the returns to education, the amount of education one chooses is a function of ability, family background, pre-market choices, as well as the expected returns associated with such a level of education.

Despite attempts to incorporate individual heterogeneity into the modelling procedure, one can argue too, that within an investment approach to modelling the returns to human capital, that risk and uncertainty are integral to the valuation process of the expected returns. Risk and uncertainty affects pre-market choices as much as they affect valuations of expected returns (Willis & Rosen, 1979). Agent responses to uncertainty are determined by the amount and nature of the uncertainty as well as by tastes and attitudes of agents in the economy (Becker, 1964).

When valuing the future payoffs associated with educational choices, one must consider the distribution of income associated with those choices (Friedman, 1953). The choice set will then be determined by the individual agent's taste for risk in the economy. Differences in tastes will dictate different choices under the same alternatives (Friedman, 1953).

An application of this can be found in Becker's (1957) work in which he models the expected race differential in wage returns from attending college. He models a labour supply curve that relates the fraction going to college with the gain expected from college. It is assumed that these curves are positively inclined. The elasticity of labour supply is determined by the dispersion around the average in ability, availability of financing, tastes and attitudes towards risk (Becker, 1957). He further notes that comparing the investments of heterogenous groups is interesting because the private rate of return for each group would be calculated using the risk profiles per individual or group (Becker, 1975).

Working within the human capital framework, Mincer's (1970) model ascribes returns to individual characteristics such as demographic factors as well as job characteristics. The calculation of expected returns from schooling are decomposed into marginal returns from each of these individual characteristics, as well as the marginal returns from job type. Theoretically, this model can be adapted for individual risk preferences by using different discount factors in the valuation of returns. However, empirical literature on modelling the returns to educational choices have not assessed the risk profiles on individuals making education choices. Rather, much of the literature linking risk and the return to human capital has been concerned with the returns to riskier jobs, given a certain education level (eg. Garen, 1988). In part, this can be attributed to the lack of objective measures of risk in evaluating peoples preferences for certain levels of education investment.

It is important, within the context of this study, to define risk. Common to existing literature on risk, risk taking refers to constructs such as goals, values, options and outcomes (cited by Byrnes, Miller, Schafer, 1990). Within this context, risk refers to taking an option when the potential outcome is not strictly positive. Building on this, Furby and Beyth-Marom (1992) define the act of implementing a goal directed option as "a risk-taking behaviour" if it satisfies two conditions. The first necessitates that the behaviour in question could lead to more than one outcome and secondly, at least one of the outcomes is undesirable. The breadth of this definition of risk allows one to classify the event "attempting a multiple-choice question with a penalty for an incorrect answer" as behaviour consistent with risk taking.

Women or students from previously disadvantaged home environments might display a greater sense of risk aversion than other male students or more privileged students. This may be a factor of the home environment that students grew up with. An example is girls from disadvantaged home environments face a higher level responsibility within the home. Their role within the home is seen as a nurture role, whereas men do not have this level of responsibility. For many girls in higher education in developing countries, their pursuit of a higher education run in discontent to family values and as a result, they display a greater sense of risk aversion.

2.1 Empirical results of the gender differential in academic performance

In the education economics literature, males typically perform better than females in economics (Siegfried, 1979). Siegfried (1979) hypothesised this was due to the cultural environment in which girls grow up as well as chromosome-linked to the difference in male and female performance in learning spatial and numerical skills, which are critical to the understanding of economics.

A socio-cultural explanation for this gender differential in performance predicts that in some cultures, women are discouraged from engaging in so-called male-dominated courses (Heath, 1989). Since course selection is voluntary, female students who choose to study business related, or "male-dominated" courses, are inherently more analytical and suited to these courses. If this explanation were true, then the gender gap is actually underestimated because of female self-selection out of business-related courses.

In the economics literature, Heath (1989) finds evidence that supports a sustained male advantage when assessment is through the multiple-choice format. Using results from the Scholastic Achievement Test (SAT), Byrnes and Takahira (1993) find evidence of a gender differential in favour of males (cited by Byrne, Miller and Schafer, 1999). Similar results have been observed by Blau, Ferber and Winkler (1998) in the SAT and by Paglin and Rufolo (1990) (cited by Altonji & Blank, 1999). Typically, this differential has been explained by coursework and maths knowledge. Altonji and Blank (1999) note that they suspect that this gender gap is due, in part, to pre-market selection.

In spite of this empirical evidence, the hypothesis that male students systematically outperform females is not supported in all the literature. Heath (1989) proposes that female students outperform their male counterparts when the assessment is based on verbal reasoning skills. Question design and content pedagogy have been used to explain differential performance in assessment in the education and psychology literature. In essay type assessment, it was found by Lumsden and Scott (1987) that women score equally or better than men when the testing format involves verbal reasoning (cited by Heath 1989).

With regards to the multiple-choice format, test structures that penalise for guessing have been studied extensively. Sherrifs (1954) find evidence that guessing is discouraged when test formats include penalties for guessing. In these test formats, there is a higher incidence of incomplete answers, even on questions that students have shown an understanding of in other testing formats. This tendency was confirmed by Ben-Shakhar and Sinai (1991) who further find evidence that females display a high rate non-response as opposed to guessing.

This paper argues that differentials in performance in multiple-choice formats can be explained by individual risk preferences as well as demographic factors. Demographic factors such as wealth, age, education, gender, religion and nationality influence attitudes towards risk but it is conceivable that two individuals with similar observable characteristics have different risk preferences (Halek & Eisenhauer, 2001).

The idea that women display a greater aversion to risk in competitive arenas has been explored by authors such as (Gneezy, Niederle, & Rustichini, 2003), (Lavy, 2004) and (Paserman, 2007). For example, in competitive tennis, Paserman (2007) finds evidence that women's performance deteriorates significantly as the importance of the points in a tennis match increases. No such evidence was found for men. These authors all attribute a part of the gender differential in academic performance or compensation to a gender difference in preference for the competitive environment governing the choice set. The degree of risk aversion determines this preference. In social psychology this phenomenon has been termed "choking under pressure" where performance is suboptimal despite a high degree of achievement motivation and high ability (Paserman, 2007).

3. Study Context

Since 1994, the political framework in South Africa has ensured that access to higher education has increased. This has resulted in higher rates of university admission of what Cross and Carpentier (2009) term a ‘new category of students’. These students are typically from disadvantaged backgrounds where the schooling culture is founded in a community based learning culture. Students from independent schools, or former Model C schools[1], find themselves at university facing a similar learning culture to what they have become accustomed to at their schools. The university learning culture in South Africa is founded in a meritocratic framework. Individuality and competition is encouraged, and only the most adaptable students survive (Cross & Carpentier, 2009). Students come to university from a diverse range of socio-economic backgrounds. The spectrum of students ranges from students from rural, under-resourced school, to students from highly resourced and sophisticated urban schools. The immediate result of the new political framework in South Africa is a heterogeneous student body that reflects different skills sets, academic backgrounds and levels of university preparedness.

These different backgrounds and learning experiences have an influence on attitudes towards learning as well as general university preparedness. Students have a diverse range of motivations for studying, again influenced by their community culture. It is thus necessary to include attitudes towards risk into Becker’s human capital model framework when analysing student performance in South Africa.

Increasing student numbers has increased the demands on academic staff in tertiary institutions, forcing the universities to seek out alternative methods of instruction and assessment. Computers and internet tools have provided many innovative ways to deal with the increased demands on resources. Most students in undergraduate courses in

[1] Former Model C school are schools that are historically white-only schools. From the early 1990’s these schools were allowed to admit black pupils up to a maximum of 50% of the student body. The term or the law no longer remains. In general, however, these schools are still better resourced and managed than other schools. This is likely to be due to greater involvement of governing bodies comprised of parents of students in the school.

South African university business faculties are assessed using a mixture of multiple-choice and short answer type formats. The immediate advantage of the multiple-choice format is that it allows for prompt grading and immediate feedback for the students.

4. Data

The research takes the form of an econometric study of a sample of 1554 students who wrote the mid-year examination for the introductory micro-economics course. This is a first year course but there are students who are in their second year of study but failed in their first year and are repeating this course to gain the credit.

The sample contains 753 males and 801 females, while approximately 61 percent of the sample is of black ethnicity. The rest of the sample is white, coloured[2] or Asian (of Indian or Eastern descent). The ethnic composition of the sample is shown below in Table 1. The students are drawn from a multitude of disciplines, including the arts, business faculties and the sciences. The degree choice is used in the analysis in order to control for motivation. Some students have chosen economics (sciences, arts) while others are required to complete the unit for their degrees (accounting, business faculties, building sciences).

Table 1: Ethnic composition of sample

	Total (N)	Male (%)	Female (%)
Sample	1554	48.43	51.54**
Black	941	45.06	54.94***
White	257	64.59	35.41***
Coloured	46	39.13	60.87**
Asian	310	46.77	53.23*

Notes: Mean for female significantly different from mean for male
* p<0.1; ** p<0.05; *** p<0.01

The unique nature of this dataset allows one to assess the academic performance differential while stratifying across both gender and race. The students in this course are culturally diverse and heterogeneous in terms of home language, gender, race and university preparedness. On average, female students performed worse in this exam than male students, scoring on average 3 percent lower than male students. This is on average

[2] The term coloured is used to denote the ethnicity of descendants of the children of white settlers in South Africa and indigenous inhabitants of the region. It is a recognised ethnic group in South Africa and the group represents around 8 percent of South Africa's population.

true for black, white and coloured students. Within these ethnic groups, male students perform better than female students. White students outperformed black students for the sample as a whole. The relationship observed for the other race groups is not observed for Asian male and female students. Table 2 below shows the average mark obtained in this exam stratified by race and gender.

Table 2: Average mark by race and gender

	Total (%)	Male (%)	Female (%)
Sample	39.97	41.36	38.66***
Black	38.11	39.59	36.89**
White	46.79	47.85	44.87*
Coloured	41.05	43.56	39.43
Asian	39.79	38.81	40.64

Notes: Mean for female significantly different from mean for male * p<0.1; ** p<0.05; *** p<0.01

The exam was comprised of 60 multiple-choice questions with five alternatives. The exam mark is calculated with negative marking. A correct answer received 4 points, 1 point was deducted for every incorrect answer and no point allocated for a blank answer. This marking scheme has an influence on whether or not a student answers a question since there is a perception of a negative payoff, even though the expected payoff from a guess is zero.

The rate of leaving out a question varies by gender and ethnicity. Males are less likely to leave out a question than female students. Black students as well as students of Chinese and Indian descent are, on average, more likely to leave out questions than white students. The average leave out rate, stratified by gender and ethnicity is shown in Table 3 below. In all, the dataset contains 93 240 data points (1554 students, each with 60 questions to attempt).

Table 3: Average rate of leaving out a question by race and gender

	Total (%)	Male (%)	Female (%)
Sample	19.47	18.39	20.49*
Black	19.11	17.55	20.38*
White	18.48	17.70	19.89
Coloured	18.91	15.93	20.83
Asian	21.49	21.92	21.10

Notes: Mean for female significantly different from mean for male * p<0.1; ** p<0.05; *** p<0.01

Students were surveyed on their attitudes towards risk. The survey format as well as the calculation of the risk measure are described in the methodology section below. Within this sample, 41 percent of females in the sample are labelled risk averse, while only around 38 percent of the males are given the same designation. This proportion differs by ethnicity, with black students reporting the higher rate of risk aversion, as shown in Table 4 below.

Table 4: Average reporting of risk aversion by race and gender

	Total (%)	Male (%)	Female (%)
Sample	39.54	37.78	41.19*
Black	42.81	41.78	43.65
White	28.57	27.54	30.43
Coloured	32.61	33.33	32.14
Asian	39.74	38.36	40.96

Notes: Mean for female significantly different from mean for male * p<0.1; ** p<0.05; *** p<0.01

Female students report a higher rate of risk aversion on average, as well as demonstrate a higher rate of leaving out questions. Within this testing format, female students perform worse than male students on average, with white students performing better than black students.

5. Methodology

The method in which students attempt and answer questions in a multiple-choice framework with negative marking is not random. Only students who responded favourably to the multiple-choice questions format are able to adapt and perform in the course. Students hold some general preferences for risk that may be thought of as time invariant. In addition, the structure of the testing format dictates that students who are less risk averse are more inclined to attempt the question. Thus, a student's level of risk aversion is also affected by the natural progression of the exam itself.

In defining the time invariant risk measure, $risk_i$, the students were asked a series of questions adapted from Eckel & Grossman (2008). Two questions were used in this study to measure risk aversion. Both frameworks do not include the possibility of losses. Each gamble included one sure payoff option. The first uses an investment framework, or rather the multiple-choice framework that the students are familiar with. The second uses an abstract gambling framework where expected payoff increases linearly with risk. These measures are unable to predict risk seeking behaviour, only attitudes towards risk aversion. In addition, this risk measure is fixed for each student. Both of these question formats are described in Appendix 1.

A student's question-specific risk can be associated with the importance attached to answering that specific question. This importance changes as the test progresses because students adapt their behaviour dynamically in this testing format. Students appear to respond to changes in the structure of the available alternatives and to the presence of time pressure (Payne, Bettman, & Johnson, 1993). As students omit questions, their expected probability of passing the exam decreases. This, in turn, has two consequences: (i) the chance of guessing an alternative in later questions increases, thereby decreasing the leave out rate in later questions, and (ii) the chance of returning to earlier questions and attempting them increases. In other words, as a student leaves out more questions, the student may adapt his or her behaviour in order to answer as many questions as possible to pass the exam. The importance attached to each question can be calculated as follows:

$$importance_{i,t} = \Pr(Pass\ exam\ as\ a\ result\ of\ answering\ q_t\ correctly \mid k\ questions\ already\ left\ out) - \Pr(Pass\ exam\ as\ a\ result\ of\ answering\ q_t\ incorrectly \mid k\ questions\ already\ left\ out)$$

There is no way to determine the order in which students answered the exam questions since the students manually filled out the computer cards while completing the exam. This paper simplifies the dynamics by assuming that students do not return to previous questions. In addition, students made a decision to attempt question t given that they have made the same decision for questions 1 to $t-1$. Importance for each question increases non-linearly throughout the exam as the student omits more questions and decreases as the student attempts more questions.

One possible mathematical relationship[3] that reflects the way importance evolves over time is:

$$importance_{i,t} = \frac{(cumulative\ number\ left\ out)^2{}_{i,t-1}}{(number\ attempted)_{i,t-1}}$$

where $importance_{i,t} \in [0,1]$. This scaling[4] is necessary in order to have comparable importance measures across different students. Also, the importance attached to question t is a function of all the questions from 1 to $t-1$.

The estimation strategy can be modelled by three separate equations as follows:

[3] An actual measure could be computed using the probabilities above. However, the fact that students could return to any question at any time during the exam implies that a probability calculated this way would not be entirely correct either. The measure calculated here is a simplification takes into account the a priori relationship between exam progression, answers already left out and answers attempted.

[4] $new\ importance_{i,t} = \frac{importance_{i,t}}{\max[importance_{i,t}]}$ for person i.

(1)

$$\Pr(attempt_{i,t}) = \kappa + \beta_1 female_i + \beta_2 risk_i + \beta_3 female_i * risk_i + \beta_4 importance_{i,t} + \beta_5 importance_{i,t} * risk_i + \beta_6 x_{it} + c_i + \xi_{it}$$

(2)

$$\Pr(incorrect_{i,t}) = \alpha + \varphi_1 female_i + \varphi_2 risk_i + \varphi_3 female_i * risk_i + \varphi_4 x_{it} + c_i + \varepsilon_{it}$$

(3)

$$\Pr(incorrect_{i,t} | attempt_{i,t}) = \alpha + \gamma_1 female_i + \gamma_2 risk_i + \gamma_3 female_i * risk_i + \gamma_4 x_{it} + c_i + \varepsilon_{it}$$

x_{it} is a vector of strictly exogenous explanatory variables, other than gender and risk preference. These include ethnicity, age, degree choice, Grade 12 aggregate mark and whether a not the student is supported by financial aid. ε_{it} and ξ_{it} are unobserved disturbances. They are assumed independent across individuals but freely correlated across time, for a given individual, i. c_i is a vector of unobservable characteristics.

The dichotomous nature of the dependent variable necessitates a non-linear estimation model such as a probit model. The equations are estimated using a standard probit model with cluster corrected standard errors rather than a population averaged probit model because the latter requires exchangeable correlation between individuals. This is a strong assumption and is not evidently true from looking at the data. The two estimators would produce the same estimates but the population averaged probit model would produce smaller standard errors.

There may be some autocorrelation in the way students answer questions, with student behaviour on question t being a function on that student's behaviour on questions 1 to $t-1$. This will be controlled for with the question-specific risk measure, $importance_{i,t}$. This measure is a function of student i's cumulative behaviour on the exam.

Using conditional probabilities and dropping the subscripts for ease of exposition, it is noted that:

$$\Pr(incorrect \mid attempt) = \frac{\Pr\,(incorrect\ \cap\ attempt)}{\Pr\,(attempt)}$$

If one assumes that the events "answering a question incorrectly" and "attempting a question" are independent, then one can write:

$$\Pr(incorrect \mid attempt) = \Pr\,(\text{incorrect})$$

This paper will test the hypothesis

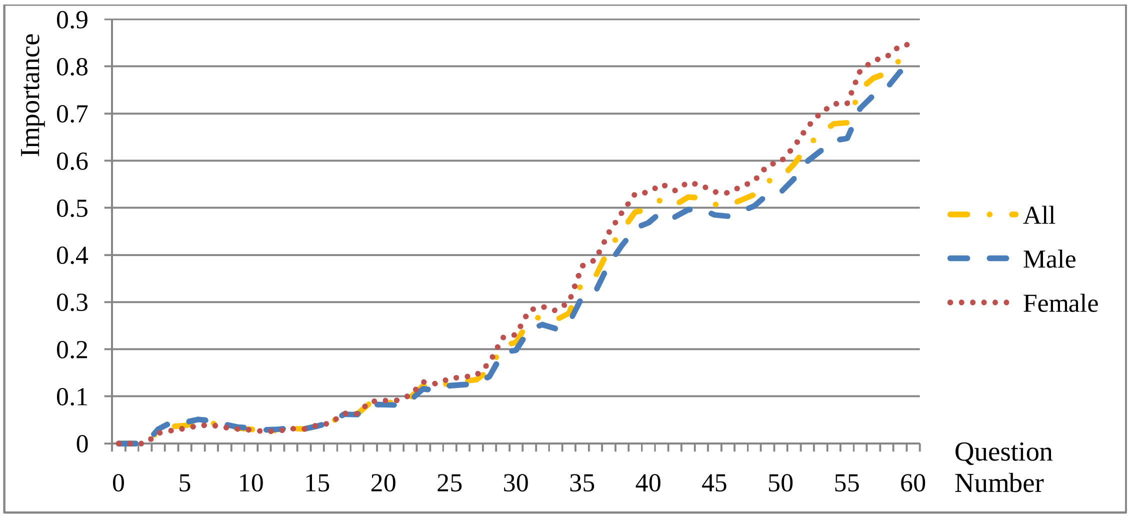
$$H_0: \gamma_1 = \varphi_1$$
$$H_1: \gamma_1 \neq \varphi_1$$

If γ_1 in (3) is significantly different φ_1 in (2), it must be that the events are not independent. The difference can be attributed to the probability of actually attempting the question in the first place, which could be determined by a student's level of risk aversion.

6. Results

As a student progresses through the test, the importance attached to each question increases. By construction, this variable increases as students leave out more questions and declines when a student attempts an answer. The importance attached to each question in the exam is different, on average, for males and females. Females have a higher rate of leaving out questions, so the attached importance increases faster for females than it does for males. This relationship is shown in Figure 1 below. Compared to the average for the sample, female importance per question increases faster than it does for males.

Figure 1: Calculated importance by question number by gender



In order to control for student ability, the students' high school year 12 aggregate mark is included in the estimation. Students are accepted into university based on their year 12 marks. The exams students write in year 12 are standardised and coordinated by Umalusi, a government department tasked with education quality assurance and accreditation. The year 12 mark enters into the estimation in a natural logarithmic form.

Also included into the regression is a variable that indicates whether or not the student has failed the course before and is repeating it to earn the credit. Whether or not the student is supported by financial aid is also included in the estimation. While controlling for motivation, the latter might be thought of as a control variable for socio-economic background.

Since the data has been pooled and estimated using a probit model, the standard errors are calculated taking into account that observations occur in clusters. Each individual is repeated for each question in the test. One cannot assume that within a cluster, the questions are uncorrelated. There is not equal correlation between individuals. Students respond to question difficulty in a non-uniform way depending on ability and understanding of the tested topic.

The results for equations (2) and (3) are shown in Table 5 below. The second and fourth columns include a gender-race interaction term, black*female. Equation 2 is for the full sample of questions. Females are 4 percent more likely to answer a question incorrectly. For questions that were attempted (Equation 3), females are approximately 9 percent more likely to get the question wrong. Higher question importance is associated with a higher probability of getting the question wrong, but for equation 3, one cannot determine whether that was because of the question importance or whether it was attempted or not. For the restricted sample, equation 3, the point estimate on question importance is not statistically significant.

Black students are approximately 11 percent more likely to get the question incorrect, both in the restricted sample and the unrestricted sample. These results are significant at the 1% and 5% significance levels for the unrestricted and restricted samples respectively. The point estimate for the interactive term, black*female, is not significantly different from zero.

Table 5: Estimation results for Equation 2 and Equation 3

Dependent variable	Equation 2		Equation 3	
Incorrect	(a)	(b)	(a)	(b)
Independent variables				
Female	0.0441*	0.0456	0.0958***	0.0842**
	(1.89)	(1.40)	(3.69)	(2.31)
Risk	-0.0087	-0.0088	0.0141	0.0149
	(-0.29)	(-0.30)	(0.44)	(0.47)
Female * Risk Averse	-0.0043	-0.0041	-0.0315	-0.0329
	(-0.12)	(-0.11)	(-0.79)	(-0.82)
Importance	0.2039***	0.2038***	0.0577	0.0579
	(4.85)	(4.85)	(1.49)	(1.49)
Importance * Risk Averse	-0.0657*	-0.0657*	-0.0523	-0.0528
	(-1.74)	(-1.74)	(-1.36)	(-1.37)
Black	0.1198***	0.1209***	0.1151**	0.1065**
	(2.77)	(2.65)	(2.39)	(2.09)
Black * Female		-0.0026		0.0194
		(-0.07)		(0.47)
Age	-0.0065	-0.0065	-0.0214**	-0.0212**
Coloured	0.0916	0.0911	0.0722	0.0753
Asian	0.1234***	0.1231***	0.1869***	0.1894***
English at home	-0.0522	-0.0521	-0.0788*	-0.0797*
Ln (Year 12 Aggregate)	-4.7747***	-4.7775***	-7.6456***	-7.6231***
Residence	0.0197	-0.0197	-0.0016	-0.0017
Repeat	-0.2109***	-0.2108***	-0.2312***	-0.2316***
Financial Aid	0.0057	0.0057	0.0185	0.0191
Other control variables				
Degree choice	Yes(***)	Yes(**)	Yes(***)	Yes(***)
Question number	Yes (***)	Yes(***)	Yes (***)	Yes(***)
Constant	7.1623***	7.1663***	11.7917***	11.7591***
Pseudo R2	0.0772	0.0772	0.1283	0.1283
N	93240	93240	75052	75052
Number of clusters	1554	1554	1554	1554

Notes:
(i) * p<0.1; ** p<0.05; *** p<0.01
(ii) t-values reported in parenthesis for selected point estimates.

In order to test the key hypothesis of this paper, one may not use the classic form of the Hausman test since it assumes that one is using the most efficient estimator (Hausman, 1984). In the presence of clustered standard errors, one should use a more general form of the test in order to assess if there is a difference between the point estimate in the restricted regression and the regression on the full sample (Weesie, 1999). Please see the appendix for the derivation of this test.

Taking into account that the data is arranged in clusters, a generalised Hausman-type test that incorporates the simultaneous covariance between the two estimators reveals that the one can reject the cross-model hypothesis that $\gamma_1 = \varphi_1$. In other words, the coefficient on female in the unrestricted equation (equation 2) is significantly different (at the 5% significance level) from the coefficient on female in the restricted equation (equation 3). The Wald-type test statistic for H_0 is H = 5.14 which follows asymptotically a χ^2 distribution[5] [6].

A similar test on the point estimates on the ethnicity variable, Black, reveals that the Wald-type test statistic for H_0 is H = 0.34 which follows asymptotically a χ^2 distribution[7]. One cannot reject the hypothesis that the point estimate is the same in the two estimations. One cannot reject the hypothesis that the point estimate of black female is the same in each equation.

Testing the joint hypothesis that all three point estimates (female, black and black*female) are significantly different from each other in each equation shows that one may reject this hypothesis. The Walt-type estimate for H_0 is H = 17.73 [8].

From these results, there is evidence that there is not independence between answering a question and getting it wrong for females. There is heterogeneity in the way students attempt the question in the first place. This suggests that academic performance is a function of pre-market selection. This selection is determined by individual characteristics.

---

[5] (Prob > chi2 = 0.0234)

[6] The tests were performed comparing Equation 2a with Equation 3a and Equation 2b with Equation 3b. The test statistics reported here are for the ‘b’ equations. The same conclusions were reached when comparing the ‘a’ equations.

[7] (Prob > chi2 = 0.562)

[8] (Prob > chi2 = 0.000)

Table 6 below shows the regression output from estimating the probability of answering a specific question (equation 1). The first column is a standard probit regression, while the second, third and fourth columns take into account the fact that the observations are clustered by individual student. The standard errors increase with this clustering. The clustered standard errors reduce the significance of the risk aversion dummy variable. Even while including dummy variables for each question, the importance variable is significant at the 1 percent level. As the question increases in importance, students are more likely to attempt it. Controlling for risk aversion and the question specific importance, females are 10 percent less likely than male students to attempt the question.

In column 3, an interactive term between the calculated question importance and whether or not the student is female is included in the regression. Since $importance_{it}$ is between 0 and 1, one must interpret such a coefficient carefully. Moving from a question of zero importance to a question of absolute importance, a female is 24 percent more likely to answer it. However, each subsequent question does not change in importance by such a large differential.

Black students are no more likely than other students to leave out questions. The inclusion of an interaction variable for black females reduces the magnitude of the point estimate on females alone. Black females are more likely to omit an answer compared to other students, although this result is not significant.

One can interpret these results by inferring that a female is 19 percent, at an upper bound, less likely to answer a multiple question than a male, for a constant level of question importance. Females are marginally more likely to answer a question of higher importance compared to one with a lower importance. Black students are no more likely to attempt questions than other students.

Table 6: Estimation results for Equation 1

Dependent Variable	Equation 1			
Attempt	(1)	(2)	(3)	(4)
Independent Variables				
Female	-0.1065***	-0.1065***	-0.1915***	-0.1566***
	(-7.78)	(-3.23)	(-4.90)	(-3.21)
Risk	-0.0607***	-0.0607	-0.0589	-0.0612
	(-3.23)	(-1.38)	(-1.34)	(-1.39)
Female * Risk Averse	0.0594***	0.0594	0.0610	0.0646
	(2.80)	(1.16)	(1.19)	(1.26)
Importance	0.4213***	0.4213***	0.3006***	0.2998***
	(14.84)	(7.21)	(4.79)	(4.78)
Importance* Risk Averse	-0.0509	-0.0509	-0.0588	-0.0575
	(-1.60)	(-1.01)	(-1.17)	(-1.15)
Female * Importance			0.2422***	0.2428***
			(4.93)	(4.95)
Black	0.0533**	0.0533	0.0509	0.0792
	(1.99)	(0.85)	(0.81)	(1.18)
Black*Female				-0.0601
				(-1.16)
Age	0.0328***	0.03278***	0.0319**	0.0312**
Coloured	0.0476	0.0476	0.0455	0.0358
Asian	-0.1314***	-0.1314***	-0.1312***	-0.1379***
English at home	0.0672***	0.0672	0.0655	0.0682
Ln (Year 12 Aggregate)	5.2837***	5.2837***	5.2727***	5.2127***
Residence	0.0702***	0.0702**	0.0703**	0.0703**
Repeat	-0.0532***	-0.0532*	-0.0519	-0.0507
Financial Aid	-0.0242	-0.0242	-0.0254	-0.0269
Other control variables				
Degree choice	Yes (***)	Yes (**)	Yes(**)	Yes(**)
Question number	Yes (***)	Yes (***)	Yes(***)	Yes(***)
Constant	-7.0169***	-7.0169***	-6.9381***	-6.8508***
Pseudo R2	0.1566	0.1566	0.1572	0.1573
N	93240	93240	93240	93240
Number of clusters	-	1554	1554	1554

Notes:

(i) * p<0.1; ** p<0.05; *** p<0.01

(ii) Column (1) reports point estimates from a standard probit model. Columns (2), (3) and (4) are point estimated from a probit model with standard errors calculated by taking into account the clustering by individual student. Column (3) augments the probit estimation with the addition of an interaction variable (female*importance), while column (4) adds (black*female) to the estimation.

(iii) t-values reported in parenthesis for selected point estimates.

7. Discussion

The results discussed above display evidence that women are less likely to attempt multiple-choice answers in the presence of negative marking. The coefficient on time-invariant risk, however, is not conclusive. Using the importance variable, females are more likely to answer a question that has a higher level of calculated importance. Feeding this into the restricted regression, females are more likely to get the answer wrong, given that they have attempted it. This supports the work of Paserman (2007) who found that women are more likely to make unforced errors as the stakes of a tennis match increase.

Black females are moderately less likely to attempt a question and if they do, are no less likely to get it wrong compared to females of the other ethnicities. There is evidence to suggest that female students are more likely to omit questions but the results do not attribute it to the risk variable included in the model. There is scope for further research to use different measures of risk, as well as other testing techniques to assess whether there is a relationship between gender biased risk preferences and academic performance. Supporting the work of Altonji and Blank (1999), these results may not hold if the testing is in another format, such as evaluation based on essay-based verbal reasoning.

Black students are no less likely to attempt answers but are more likely to get the question incorrect. This may reflect prior learning experience and a lower level of university preparedness as a result of differential school quality. Further research could include better controls for school quality such as region and income quantile of the school district.

These results suggest that schools and tertiary education institutions should be cognisant that their testing formats might not be testing knowledge. Instead they may be disadvantaging some students based on prior schooling experience.

The implications for the labour market are far-reaching. The results have implications on gender differences in pre-market selection. Women may earn less than their male counterparts because they self-select themselves into lower paying jobs that have greater job security. Further, these results provide additional support for a signalling model similar to Spence (1973). Risk aversion may define an individual's opportunity set. The opportunity costs of investment into human capital may be different and thus different levels of education may signal ability to potential employers. If risk aversion affects these decisions, then there may be dissimilar outcomes for men and women, given a certain level of ability. As an illustration, on advising a female client, an insurance salesman may offer his client products tailored for generally riskier clients. Females end up paying more for insurance than they would otherwise pay if they were offered another menu of options. In the same way, employees may offer female workers lower wages than men.

In order to test this, one would need to incorporate better control variables for risk aversion. This paper uses a simple dummy variable derived from the answer to two questions assessing risk aversion in order to calculate a time-invariant risk measure. A more composite measurement scale, perhaps drawn from psychology literature, could enhance the analysis. The same could be said for the time-varying risk measure.

The issues that this paper raises suggest that if steps were taken to address the preferences of students, one might be able to have a small effect on reducing inequality between males and females. For women to be able to compete effectively with men in the classroom setting, and by implication in the workplace, it is necessary both to reduce the risk aversion of females as well as change the employer's perceptions of females being more risk averse.

The human capital model uses the concept of expected returns to model optimal investment in human capital. When expected returns are a function of individual heterogeneity, there results in a wide set of investment choices. Scope for further research lies in understanding the pre-market selection into these investment choices, both as a function of the expected returns as well as of the determinants affecting the valuation of those returns.

8. Conclusion

This paper has shown evidence that student performance is not independent of pre-market selection. There is a gender differential in performance, but part of this differential can be explained by the different strategies student adopt as they progress through an exam. Controlling for individual importance attached to each questions, females are at an upper bound, 19 percent less likely to answer multiple questions that have an inbuilt penalty for being incorrect. A similar result was not found for black students, even though there remains a race differential in academic performance. This result for black students is surprising, since the mean academic performance of white students across genders is different from the mean academic performance for black students.

With regards to academic performance, female students are approximately 9 percent more likely to get an answer incorrect, given that they have attempted the question. Black students are approximately 11 percent more likely to get an answer incorrect, given that it was attempted.

Testing format appears to disadvantage some students more than others. While this paper has been unable to attribute this to risk, there is evidence that it is as a result of the way students learn in different contexts and approach specific assessment formats. There is scope to replicate this study using alternative assessment methods and in other subject areas.

Appendix

Appendix 1 – Calculation of t-statistics for difference in means

Table 1: Ethnic composition of sample

	Male			Female			
	mean	se	n	mean	se	n	t
Sample	48.43	0.4998	753	51.57	0.4997	801	-1.24
Black	45.03	0.4975	454	54.97	0.4975	517	-3.11
White	64.48	0.4786	166	35.52	0.4785	91	4.64
Coloured	39.13	0.4881	18	60.87	0.4881	28	-1.47
Asian	46.79	0.4989	145	53.21	0.4989	165	-1.13

Note: se is calculated as the overall (within and between) standard error in the sample.

Table 2: Average mark by race and gender

	Male			Female			
	mean	se	n	mean	se	n	t
Sample	41.36	0.1798	753	38.66	0.1683	801	3.05
Black	39.59	0.1662	454	36.89	0.1553	517	2.60
White	47.85	0.1888	166	44.87	0.2035	91	1.15
Coloured	43.56	0.1698	18	39.43	0.1933	28	0.76
Asian	38.81	0.1904	145	40.64	0.1711	165	-0.89

Note: se is calculated as the overall (within and between) standard error in the sample.

Table 3: Average rate of leaving out a question by race and gender

	Male			Female			
	mean	se	n	mean	se	n	t
Sample	18.39	0.3874	753	20.49	0.4036	801	-1.05
Black	17.55	0.3805	454	20.38	0.4028	517	-1.13
White	17.70	0.3817	166	19.89	0.3992	91	-0.43
Coloured	15.93	0.3661	18	20.83	0.4062	28	-0.42
Asian	21.92	0.4137	145	21.10	0.4081	165	0.18

Note: se is calculated as the overall (within and between) standard error in the sample.

Table 4: Average reporting of risk aversion by race and gender

	Male			Female			
	mean	se	n	mean	se	n	t
Sample	37.78	0.4848	753	41.19	0.4922	801	-1.38
Black	41.78	0.4932	454	43.65	0.4959	517	-0.59
White	27.54	0.4467	166	30.43	0.4601	91	-0.49
Coloured	33.33	0.4716	18	32.14	0.4601	28	0.08
Asian	38.36	0.4862	145	40.96	0.4917	165	-0.47

Note: se is calculated as the overall (within and between) standard error in the sample.

Appendix 2 – The survey to measure risk aversion

The calculation of the $risk_{it}$ dummy variable uses the two questions listed below. To be classified as risk averse (i.e. $risk_{it} = 1$), the student had to answer option A for both questions.

Question 1

Imagine the marking allocation for a multiple-choice question for FIVE options is as follows: correct answer = 4 marks, incorrect answer = -1 (i.e. negative marking), and no answer = 0 marks. Which of the following applies to you?

A. You will only answer the question if you are 100% sure if the answer is correct.
B. If you believe that the right answer is one of TWO options (i.e. you have eliminated three options as obviously incorrect, you will guess between the two options.
C. If you believe that the right answer is one of THREE options (i.e. you have eliminated two options as obviously incorrect, you will guess between the three options.
D. If you believe that the right answer is one of FOUR options (i.e. you have eliminated only one option as obviously incorrect, you will guess between the four options.
E. You will guess even if you are unable to eliminate any of the options as obviously incorrect.

Question 2

You are invited to take part in a money game which works as follows: You can choose to play in one of five games, which are called A, B, C, D and E. Once you have chosen a game to play, a coin is thrown to determine how much you will win. The games are the following:

Game	Winning amount if heads	Winning amount if tails	Expected value of game	Variation in payoffs (risk)
A	R 50.00	R 50.00	R 50.00	R 0.00
B	R 75.00	R 37.50	R 56.25	R 37.50
C	R100.00	R 25.00	R 62.50	R 75.00
D	R125.00	R 12.50	R 68.75	R112.50
E	R150.00	R 0.00	R 75.00	R150.00

Appendix 3 – derivation of the generalised Hausman-type test in the presence of clustered standard errors

This derivation is taken from Weesie's (1999) note on derivation of the Wald-type test statistic in the presence of clustered standard errors.

Since we have estimated the model on the same data, where one model is run on a subset of the full sample, it can assumed that the estimators are stochastically dependent. To derive a test for the cross-estimator hypothesis, one requires a measure of the cross sample variance V_j which is found by using an estimate of the covariance of the different estimators, V_{jh}.

In the standard Hausman (1978) case, with k=2 and an equality hypothesis
$H_0 : plim\ b_j - b_h = 0$, we have

$$V = AVar(b_j - b_h) = V_j + V_h - V_{jh} - V'_{jh}$$

The Wald-type estimate can be calculated by

$$H = (b_j - b_h)' V^{-1} (b_j - b_h)$$

Under H_0, this test statistic follows a χ^2 distribution with K degrees of freedom. We reject H_0 if H is large, relative to the χ^2-derived threshold.

For the generalised form of the Hausman-type test used in this paper, one does not require the efficiency assumption as in (Hausman J. A., 1978). The two equations are estimated simultaneously and clustered standard errors are calculated in order to derive V.

References

Altonji, J. G., & Blank, R. M. (1999). Race and Gender in the Labour Market. In O. C. Ashenfelder, & D. Card (Eds.), *Handbook of Labour Economics* (pp. 3143-3259).

Becker, G. S. (1957). *The Economics of Descrimination.* Chicago: The University of Chicago Press.

Becker, G. S. (1964). *Human Capital: A theoretical and empirical analysis, with special reference to education.* New York: National Bureau of Economic Research.

Becker, G. S. (1975). *Human Capital: A theoretical and empirical analysis , with special reference to education.* New York: National Bureau of Economic Research.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender Differences in Multiple Choice Tests: The Role of Differential Guessing Tendencies. *Journal of Educational Measurement*, 23-35.

Byrne, J. M., Miller, D. C., & Schafer, W. D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin, 125*(3), 367-383.

Card, D. (1999). The Causal Effect of Education on Earnings. In O. C. Ashenfelder, & D. Card (Eds.), *Handbook of Labour Economics* (Vol. 3, pp. 1801-1863). Elsevier Science.

Cross, M., & Carpentier, C. (2009). New Students in South Africa Higher Education: Institutional culture, student performance and the challenge of democratisation. *Perspectives in Education, 27*(1), 6-18.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting Risk Attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavious and Organization*, 1-17.

Eckel, C., & Grossman, P. (2008). Men, Women and Risk aversion. In C. Plott, & V. Smith, *The Handbook of Experimental Economics.* Amsterdam: Elsevier.

Friedman, M. (1953). Choice, Chance, and the Personal Distribution of Income. *The Journal of Political Economy*, 277-290.

Furby, L., & Beyth-Marom, R. (1992). Risk Taking in Adolescence: A Decision-Making Perspective. *Developmental Review, 12*, 1-44.

Garen, J. (1988). Compensating Wage Differentials and the Endogeneity of Job Riskiness. *The Review of Economics and Statistics, 70*(1), 9-16.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender differences. *Quarterly Journal of Economics*, 1049-1074.

Halek, M., & Eisenhauer, J. G. (2001). Demography and Risk Aversion. *The Journal of Risk and Insurance, 68*(1), 1-24.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271.

Hausman, J. A. (1984). Specification Tests for the Multinomial Logit Model. *Econometrica, 52*(5), 1219-1240.

Heath, J. A. (1989). An Econometric Model of the Role of Gender in Economic Education. *The American Economic Review, 79*(2), 226-230.

Lavy, V. (2004). Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment. *NBER Working Paper No. 10678*.

Mincer, J. (1970). The Distribution of Labour Incomes: A survey with special reference to the human capital model. *Journal of Economic Literature*, 1-26.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban labour Markets. *International Economic Review, 14*(3), 693-709.

Paserman, M. D. (2007). Gender Differences in Performance in Competitive Environments. *IZA Discussion Paper Series No. 2834*, 1-61.

Payne, J., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision-Maker*. Cambridge: Cambridge University Press.

Sherrifs, A. (1954). Who is Panalized for Guessing? *Journal of Educational Psychology*, 81-90.

Spence, M. (1973). Job Market Signalling. *Quarterly Journal of Economics, 87*(3), 355-374.

Weesie, J. (1999). Seemingly Unrelated Estimation and the Cluster-Adjusted Sandwich Estimator. *Stata Technical Bulletin*, *9*(52), 34-47.

Willis, R. J., & Rosen, S. (1979). Education and Self-Selection. *Journal of Political Economy, 87*(5).