Volume 13, Number 3

Printed ISSN: 1533-3604
PDF ISSN: 1533-3590

JOURNAL OF ECONOMICS AND ECONOMIC EDUCATION RESEARCH

The *Journal of Economics and Economic Education Research* is owned and published by the DreamCatchers Group, LLC. Editorial Content is controlled by the Allied Academies, a non-profit association of scholars, whose purpose is to support and encourage research and the sharing and exchange of ideas and insights throughout the world.

The *Journal of Economics and Economic Education Research* is owned and published by the DreamCatchers Group, LLC, PO Box 1708, Arden, NC 28704, USA. Those interested in communicating with the *Journal,* should contact the Executive Director of the Allied Academies at info@.alliedacademies.org.

EDITORIAL REVIEW BOARD

EDITORIAL REVIEW BOARD

TABLE OF CONTENTS

LETTER FROM THE EDITORS

The *JEEER* is dedicated to the study, research and dissemination of information pertinent to the discipline of economics, and to the improvement of methodologies and effective teaching in economics. The *Journal* bridges the gap between the theoretical discipline of economics and applied excellence relative to the teaching arts. The Academy is an affiliate of the Allied Academies, Inc., a non profit association of scholars whose purpose is to encourage and support the advancement and exchange of knowledge, understanding and teaching throughout the world.

The Editorial Board considers two types of manuscripts: first is theoretical and empirical research related to the discipline of economics. The second area is research oriented toward effective teaching methods and technologies in economics.

These manuscripts have been double blind reviewed by the Editorial Board members. The manuscripts published in this issue conform to our acceptance policy, and represent an acceptance rate of 25% or less.

We are inviting papers for future editions of the *Journal* and encourage you to submit your manuscripts according to the guidelines found on the Allied Academies webpage at www.alliedacademies.org.

Grady Perdue
University of Houston-Clear Lake

Martin Milkman
Murray State University

John Marcis
Coastal Carolina University

ENROLLMENT PATTERNS IN HIGHER EDUCATION

Nancy J. Burnett, University of Wisconsin-Oshkosh

ABSTRACT

This paper investigates enrollment patterns in college courses using various explanatory variables such as instructor gender, time of day and student ratings from the site RateMyProfessors.com. Empirical results from this study suggest no gender preference and the expected time preferences (middle of the day and evening as desirable and an avoidance of Friday classes). Beyond these expected results, there is evidence that some of the website ratings do hold explanatory power for enrollment patterns for my university. In particular, the rating 'ease' is statistically significant while ratings for 'overall quality,' 'clarity,' and the measure of sexual attractiveness are not. (JEL A22, A11)

INTRODUCTION

Enrollment patterns have long been assumed to exist, as any current or former department chair in charge of scheduling can tell you, but have not been formally studied in the literature. Among the assumptions commonly made at this university, are timing preferences against early morning and Friday classes, for instance. Those timing preferences may vary slightly from region to region or among student cohorts, but anecdotal evidence suggests that students prefer classes in the middle of the day more than at any other time, with additional preference for evening courses, although the 'evidence' for such a pattern can only be found in *de facto* section offerings. Indeed, there are so many courses offered in the middle of the day that departments at the university under study are not allowed to create more than 10% of their sections to begin between the hours of 9:10 and 11:30, though this may reflect teaching preferences rather than student preferences. Aside from preferring these peak hours, both students and faculty may attempt to avoid Friday classes in order to facilitate longer weekends. A few students, perhaps non-traditionals or those that work full time, may have a slight preference for early sections or evening sections so they can be 'shoehorned' into already busy lives. There has been little documentation of these effects however.

This paper seeks to examine enrollment patterns exploring both timing issues as well as information regarding individual teacher characteristics, using an original data set drawn from a single, Mid-Western, State University of approximately 11,000 undergraduates. In the past, teaching characteristics other than gender have been difficult, if not nearly impossible, to obtain in any standardized way, with the possible exception of some fraternity/sorority informal ratings available only to members. With the existence, and rapid growth, of the various professor ratings

sites on the internet, ratings of individual professors have become easily available and standardized. The largest of these internet sites is www.RateMyProfessor.com.

One of the unique aspects to this website information is its ease of access to students planning course and section enrollments. Traditional student survey results are difficult to access and few students are even aware of how to do so. Indeed, what stimulated the idea for this project was the overwhelming student response to a pre-class discussion on how students choose classes/instructors. An overwhelming majority of students reported that they "always" or "often" checked www.Ratemyprofessors.com (henceforth referred to as the website) before they enrolled for classes and **not one student** reported checking on campus rating results. Granted this 'evidence' is very anecdotal and in no way can be construed as proof of student behavior, the results of this particular pre-class discussion caused this researcher to begin considering carefully the power of the website on class enrollments

The information from this website is far easier to obtain and has fewer (and more direct) categories than our on-campus surveys. Furthermore, these ratings are completely uniform, unlike the college- and even discipline- specific surveys used in some instances at our University. Each teacher has a summary entry where the teacher's department, number of ratings and composite averages from all raters can be found for 'Overall Quality' 'Ease' and an indicator variable for sexual attractiveness listed as 'Hot' which indicates if *any* raters rated this teacher as attractive. Linking further into the website provides a full entry for each teacher than includes individual ratings that show written comments from each rater as well as individual ratings for each of the categories found at this website which include 'Clarity', 'Easiness', 'Helpfulness', and 'Overall Quality' and a measure of 'Hotness' (which I note is not a category found in any of the traditional surveys here on campus) as well as the course which is being reviewed, the date of the rating, and the rater's reported interest level before attending class. The ratings reported on the website summary for each professor give averages on a 0-5 scale 'Overall Quality' and 'Ease', along with the attractiveness variable they annotate with a Chili Pepper. When raters are creating an individual rating, however, the categories are 'Clarity' (measured from Incomprehensible to Crystal), 'Easiness' (from Hard to Easy), 'Helpfulness' (from Useless to Extremely Helpful) as well as the rater's interest level in the course before attending the class (from None to Its My World) and Textbook Usage (from Low to High). Raters are then asked whether they consider the instructor 'Hot' or not.

This website bills itself as the largest and most referenced website of its kind, specifically:

> *RateMyProfessors.com is the Internet's largest listing of collegiate professor ratings, with more than 6.8 million student-generated ratings of over 1 million professors. Each year, millions of college students use the site to help plan their class schedules and rate current and past professors on attributes such as helpfulness and clarity. Online since 1999, RateMyProfessors.com currently*

offers ratings on college and university professors from over 6,000 schools across the United States, Canada, England, Scotland and Wales with thousands of new ratings added each day ("About Us" in www.RateMyProfessor.com)

The website has a 'Terms of Use' as well as 'Site Guidelines' section that attempts to reduce fraudulent or abusive ratings. There is a moderator for ratings posts that reviews each post from a new rater (including all ratings from 'guests' to the site, though only the first review from a registered user of the site) for abusive content. Specifically:

COMMENTS THAT CONTAIN THE FOLLOWING WILL BE REMOVED :

1. *Profanity, name-calling, vulgarity or sexually explicit in nature*
2. *Derogatory remarks about the professor's religion, ethnicity or race, physical appearance, mental and physical disabilities.*
3. *References to professor's sex life (Including sexual innuendo, sexual orientation or claims that the professor sleeps with students).*
4. *Claims that the professor shows bias for or against a student or specific groups of students.*
5. *Claims that the professor has been or will be fired, suspended from their job, on probation.*
6. *Claims that the professor engages or has previously engaged in illegal activities (drug use, been incarcerated.)*
7. *Includes a link/URL to a webpage or website that does not directly pertain to the class.*
8. *Any piece of information including contact info that enables someone to identify a student.*
9. *Any piece of information about the professor that is not available on the school's website and allows someone to contact them outside of school. This also includes remarks about the professor's family and personal life.*
10. *Accusations that the professors is rating themselves or their colleagues.*
11. *Is written in a language other than English. Unless you attend a French-Canadian school*

("Rater Guidelines" in www.RateMyProfessor.com)

Posts from registered users are not automatically reviewed, but may be flagged by anyone viewing the professor's site. Once a rating is flagged, it is reviewed by the site moderators. Any posting deemed offensive is deleted from the site and removed from the composite average ratings. Further, if there are multiple ratings for a single professor from a single IP address in a

'short amount of time' they are all automatically deleted, without regard for whether said comments are positive or negative.

Additionally, if there is a threat of some kind made in any rating, the website pursues the rater as follows:

> *Comments containing a threat of violence against a person or any other remark that would tend to be seen as intimidating or intends to harm someone will (be) deleted. RateMyProfessors will notify the authorities of your IP address and the time you rated. This is enough information to identify you. IP addresses will also be turned over to the proper authorities when presented with a subpoenas or court orders from a government agency or court. ("Rater Guidelines" in www.RateMyProfessor.com)*

Even with these safeguards in place, the ratings found at this (or any other website of its kind) suffers from rather extreme response bias. Unlike university sponsored student rating surveys, website data has very undesirable properties due to the sampling process involved. Clearly only highly motivated individuals bother to enter ratings of any kind and there is no control over whether each rater has ever even taken a course from the teacher involved. Therefore, this data can only be used as a presumably exogenous information source which *may* bear no relation to actual teaching characteristics, however, whether the data is accurate or not does not affect the extent to which students may rely upon it (though, presumably, if it is found to be unreliable over time reliance upon it should diminish if users are rational). Hence, this paper investigates the extent to which this data appears to explain enrollment patterns, rather than attempt to explain the ratings themselves or their validity.

RELATED LITERATURE

Though there is little in the way of formal analysis of enrollment patterns there has been a long history of analysis of the standard forms of student evaluations of instruction. Indeed, one of the first appeared in the *Journal of Higher Education* in 1931 (Root, 1931) and presented a proto-type evaluation form and discussed how such evidence might benefit the cause of higher education. Shortly after World War II Taylor, Hankins and Lazerowitz, 1947, discussed the existence of evaluation techniques that were just then becoming common and the validity of such ratings for the analysis of teacher quality. Since that time, there have been any number of articles that appear in the top journals of many disciplines, even outside the field of education, that have discussed student evaluation ratings. Examples from economics, education, and others are cited in the bibliographies of articles referenced here, but many more exist from the fields of psychology, journalism, sociology and other disciplines.

For many years there has been great controversy over the nature, the usefulness, and the validity of student evaluations of teaching effectiveness. This debate continues in multiple forums, examining the issue from a multitude of viewpoints. Research into these questions has been done using various formats. There has been work done on experimental data (often groups of students viewing various presentations and then being asked to rate the abilities of those presenting, for instance) as well as on actual evaluation results. A good source for the historical background of these investigations can be found in McKeachie, 1990. Reliability and validity of various evaluation forms as they are commonly used in higher education (end of semester student evaluations of instruction) is discussed in Morrow, 1977.

Often the study of student evaluation results has taken the form of discipline specific analysis, such as that in Becker, 2000, in which he discusses how various classroom methodologies can play out in evaluations. Another economics-specific piece analyzes whether class size and pedagogy in introductory economics courses affects student ratings (Siegfried and Kennedy, 1995).

Much work has concentrated on what factors contribute to ratings. For instance, the issues of gender bias have arisen frequently. For instance Ferber and Huber (1975) examine student evaluations for evidence of sex bias and find that male students tend to be more critical of female instructors while female students are less harsh on female instructors. Another study (Centra and Gaubatz, 2000) finds that gender bias is significant, but often discipline specific (higher in some disciplines, such as the hard sciences, than others).

What is new here is moving away from the standard forms of student evaluation ratings that each institution offers and toward the fairly new internet version of these ratings to be found at Ratemyprofessors.com. Specifically, there have been few studies that examine the influence of an instructor's sexual attractiveness, which is now possible with data from the website (the 'chili pepper' factor).

A mere handful of years ago, sites such as this were rare, had only a sporadic following, with even more sporadic coverage by institution and discipline, much less by professor. Now this one website has millions of entries covering every conceivable discipline and thousands of institutions of higher learning. Furthermore, since these sites are in no way mandatory the raters of faculty on these sites are motivated purely by wishing to have their opinions made known as these opinions are instantly available world-wide. There is little possibility that such ratings will be traced back to any individual meaning that students have no perception that such ratings will either help or hurt their course grades. These ratings may suffer from some animosity/affinity bias that students may feel for individual teachers rather than any actual ratings of classroom performance, though the website owners claim that this site sees an average of 65% positive ratings (suggesting the site is not just a place for students to gripe). Another factor that may prove important is that ratings can be made *at any time*; during or after the student takes the class. Because students can rate professors after grades have been awarded, those ratings are

more likely to be honest about their assessments as well as more knowledgeable about whether the final grades in the course were reflective of student expectations.

METHODOLOGY

Course enrollment depends, in part, on both section specific characteristics such as time of day and general instructor characteristics (such as gender), as well as of course characteristics common to all the sections of a given course. For instance, there will be a higher demand for courses that count for general education requirements and for those that are required for a major than for those that are pure electives. Hence, explanatory variables can be partitioned into two parts; one part will be unique to each individual section while the second will be common to all courses at the university (time and day). So, we expect that enrollment for a particular course, noted as E_i will be

$$E_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon$$

where the vector *X* denotes section specific variables and the vector *Z* denotes variables common to all the sections of the same course.

The matrix of common course variables Z would be so large, relative to the number of available courses that collecting the data would be cumbersome in the extreme and the resulting model would not be identifiable.

To eliminate the common course variables (Z), the first difference between enrollments in individual sections of a course from the course average was taken, so that enrollment is standardized relative to that information. To do that, however, enrollment data needed to be comparable across sections. Therefore, each section's enrollment data was converted to a percentage of cap (so as to be comparable between courses with different course caps) and that percentage is then subtracted from the average *for that course*. There were some 64 separate courses, consisting of multiple sections, in the data set (see below). Of course, for this to provide consistent results, section caps need to be set exogenously from section demand. Of the 64 separate courses, all but 3 had very consistent section sizes. The exceptions involved two departments (Sociology and Religion) where there were 1 or more pits of 200+ and then 2 to 3 sections of 50-67 each. The number of sections involved was small enough that regression results were unaltered for signs or significance when these courses were dropped from the data set. The dependent variable, *EnrollOfAverage*, becomes

$$EnrollOfAverage_i = \frac{E_i}{cap_i} - \frac{1}{n}\Sigma_i \frac{y_i}{cap_i}$$

where cap_i is the cap for section i, E_i is the enrollment for section i and n is the number of sections for that course.

The final model then consists only of the section specific characteristics, such as professor and time, across all sections and courses.

DATA

Data was collected at the University of Wisconsin-Oshkosh for the Fall semester. This university is situated in North-Central Wisconsin and consists of slightly over 11,000 undergraduates, of whom some 59% are female, and some 15% are over the age of 25. Of the undergraduates, approximately 92% are either listed as 'white' or of 'unknown' ethnic origin. Each year we admit about 900 transfer students and about 2,000 new first year students. The instructional staff consists of 300 Faculty members and a further 268 'Instructional Staff', giving a student/teacher ratio of 21:1 with over 1,100 sections offered.

The data set consists of all multi-section, multi-instructor, lecture courses offered at the University for the Fall semester, collected across colleges (all four colleges at the University were surveyed -Letters & Science, Business, Education and Nursing - for suitable courses, though none were found in the College of Nursing. Fourteen departments from Letters & Science and six departments in Business, including Economics, and the combined introductory courses in Education are represented in the data set). All other courses, such as single section courses or those with labs or that have only one instructor with multiple sections, were excluded, leaving 306 unique sections from 64 different multiple-section courses across 20 departments. Data was collected after all continuing students' registration appointments had come and gone, so that all such students had had a chance to enroll in courses.

As of the date the data was collected, sections had not yet been dropped for low enrollment but neither had entering students (transfers and incoming freshmen) yet had a chance to register. Because of this, there are many sections that have very low enrollment in the data set, though this is customary at this point in the year, particularly those courses that traditionally are freshman classes (English Comp., Communications, and lower level Math courses for example). On the other hand, many sections had already met their caps.

The 306 sections were surveyed for current enrollment, cap, time the section is scheduled to begin, and day of week schedule. Since the University offers a rather diverse schedule encompassing not only traditional MWF or TuTh courses but also sections that meet 4 days a week, MW, and evenings once or twice per week, the data was further broken down to include indicator variables for sections beginning after 5 pm (EVENING) and whether the section has any meetings scheduled for Fridays (FRIDAY).

Additionally, as of the date the data was collected, there were still four months before the semester begins so that many classes do not yet have an instructor other than "Staff" listed. Most commonly, entry level courses in Math, English and Communications have no uniquely

identified instructor. Of the original 306 sections, 135 had 'Staff' listed as instructor. For those sections, data on time and date were available, but no instructor data. Among the remaining sections, there were only 9 named instructors that did *not* have listings on RateMyProfessor.com, leaving 145 sections with section data as well as full instructor data.

Of the 145 sections remaining in the data set, the full array of data available from the website www.RateMyProfessor.com , as well as the gender of the instructor was recorded (FEMALE, an indicator variable). Website ratings include the composite average ratings on 'Quality Overall' 'Ease' 'Helpful' and 'Clarity', each of which is fairly self explanatory as well as a rating for sexual attractiveness I have dubbed HOT and is noted on the website as a chili pepper. The first four characteristics are measured on a scale of 0-5, with 5 being highest. Attractiveness was recorded in the data set as an indicator variable. This attractiveness data is available in two forms: on the summary information per instructor there is the notation whether ANY rater gave the instructor a chili pepper (the basis for the indicator variable HOT) and on the instructor's individual page that includes written comments the total number of chili peppers awarded in total is listed. The overall number of chili peppers did not affect the models any differently than the indicator variable, so the remaining variable in the data set is the indicator variable HOT. Further, the total number of ratings a teacher received is also on the data set.

The gender of the instructor is also available from the website, as reading the comments always yielded a pronoun that determined gender. Table 1 contains variable descriptive statistics as well as Pearson correlation coefficients.

EMPIRICAL EVIDENCE

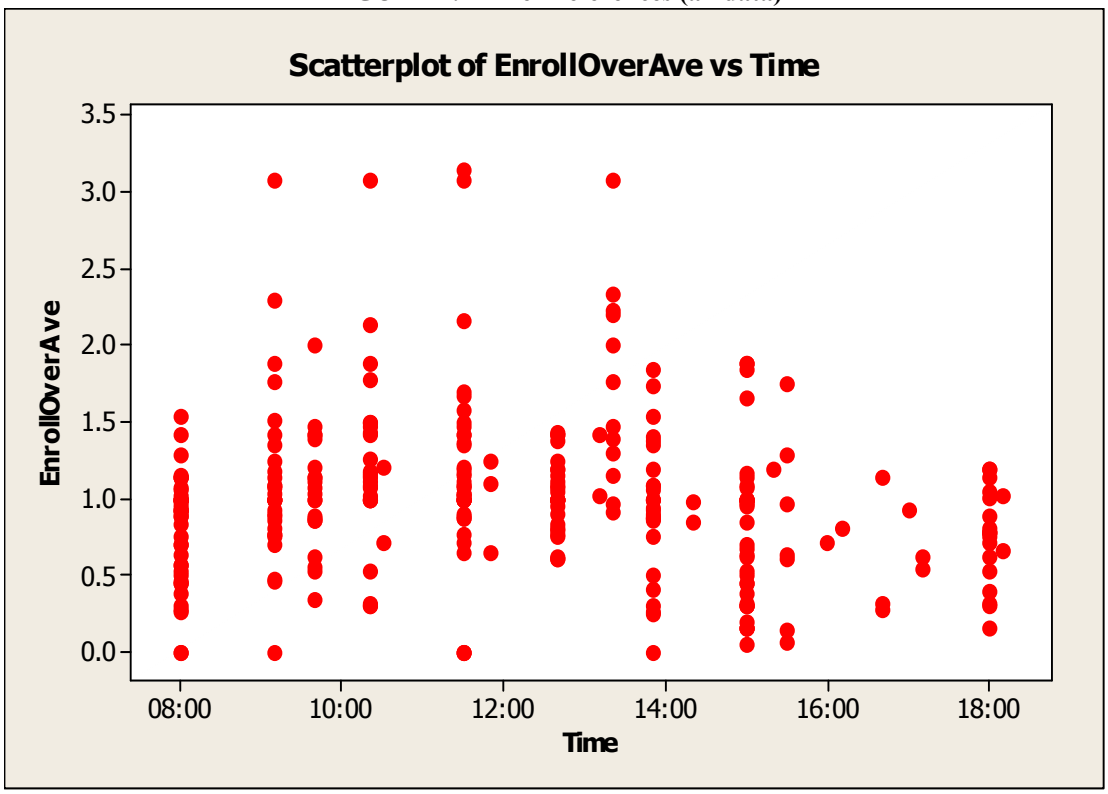
Table 1 shows there are a good array of class times and days represented, as 42.5% of the classes are MWF, 7.5% are evening classes, and good distribution of times (using military clock, coded as date/time variable). The average number of ratings for instructors is fairly high at 7.24, with a standard deviation of 10.6. Only 19.3% of instructors received one or more 'chilli peppers' ('Hot'). Looking at the correlations, we can see several variables that are significantly related, especially that data collected from the website ('Quality', 'Ease', 'Helpful', 'Clarity'), so much so that multicollinearity among these variables would swamp any direct effect. Specifically, Table 2 shows 'Quality Overall' is very nearly a linear combination of the two variables 'Helpful' and 'Clarity.' With an adjusted R^2 of .976, the RateMyProfessor.com website data is comprised of .57 * 'Helpful' and .41* 'Clarity'. The final model, therefore, omits the individual ratings of 'Helpful' and 'Clarity' in favor of the combined 'Overall' variable. Interestingly, the variable of sexual attractiveness did not show up as significant in any formulation of the overall quality rating.

Table 1: Variables-Descriptive Statistics and Correlation Matrix												
	Mean	*StD*	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
(1) EnrollOfAv	1.00	.518	**306**	306	306	306	306	145	145	145	145	145
(2) Time	.502	.121	.036	**306**	306	306	306	145	145	145	145	145
(3) Evening	.075	.264	-.058	.495*	**306**	306	306	145	145	145	145	145
(4) Friday	.425	.495	-.035	-.376*	-.197*	**306**	306	145	145	145	145	145
(5) Num.Ratings	7.24	10.6	.047	-.115	-.087	.175*	**306**	145	145	145	145	145
(6) Female	.317	.467	.008	-.111	-.084	0.00	.098	**145**	145	145	145	145
(7) Quality	3.45	.982	.180	-.033	-.031	.039	.075	.125	**145**	145	145	145
(8) Ease	3.00	.829	.214*	.015	.011	-.121	.004	-.096	.601*	**145**	145	145
(9) Hot	.193	.396	.028	-.011	-.029	-.041	.136	.305*	.381*	.003	**145**	145
(10) Helpful	3.48	1.03	.19	-.043	-.024	.069	.051	.098	.972*	.592*	.362*	**145**
(11) Clarity	3.35	.977	.163	.019	-.031	.064	.146	.117	.953*	.616*	.310*	.901*

Table 2: Explaining "Quality" **y: Quality Overall**		
	(1)	*(2)*
Constant	.071 (1.55)	-
Helpful	.572 (20.37)*	.5801 (20.91)*
Clarity	.414 (13.96)*	.4251 (14.71)*
Adj R^2	.976	
F	2991.5*	40642.66*
*significant at 5%		

Figure 1 shows the pattern of enrollment across time, leading to the conclusion of non-linearity in that pattern, so that the final models employ a quadratic term for time (TimeSquare). It appears that there is a peak during the middle part of the day (beginning around 10 am and continuing until approximately 1 pm) with generally lower enrollments for both early (those notorious 8 am sections).

FIGURE 1: Time Preferences (all data)



Using the overall data set, including the sections that list only 'Staff' for instructor limits the analysis to a small number of section specific characteristics, including time and day variables with the only instructor specific information available across sections being how may ratings each listed instructor has available for viewing on the website (with 'Staff' having no such ratings). This leads to Model (1):

$$EnrollOfAverage_i = \alpha + \beta_1 Time + \beta_2 TimeSquare + \beta_3 Evening + \beta_4 Friday + \beta_5 Num.Ratings + \varepsilon_i$$

where variables are as defined above.

This model attempts to get at whether students are enrolling in classes based on whether faculty members have large numbers of ratings on the site.

Models (2 and 3) limits the data set to those sections with identifiable instructors with at least one rating on the website so that the size of the dataset shrinks from 306 observations to 145. However, these models are able to make use of instructor data from the website. The unique characteristic ratings on the site (using the combined 'Quality' variable in place of the 'Helpful' and 'Clarity' variables) are added to the models as below:

$$EnrollOfAverage_i = \alpha + \beta_1 Time + \beta_2 TimeSquare + \beta_3 Evening + \beta_4 Friday + \beta_5 Num.Ratings + \beta_6 Female + \beta_7 Ease + \beta_8 Hot + \beta_9 Quality + \varepsilon_i$$

where variables are as already defined.

Model 2) uses the entire set of these variables while Model 3) omits 'Number of Ratings', 'Hot' and 'Quality' (variables that had weak evidence).

Table 3 shows OLS results for the three models. Residual analysis supports this choice of technique. Previous beliefs about timing preferences are confirmed. Section times are most favored in the middle of the day, with some slight additional preference for sections starting after 5 pm. There is a negative effect for sections that include a Friday meeting, though this effect diminishes once the effects for individual faculty can be accounted for. None of the models demonstrate the least significance on the gender of the instructor. As a side note, an interactive effect between 'Hot' and 'Female', labeled the 'Bimbo Effect' was also insignificant. Indeed, the only instructor characteristic that was significant was the measure of 'Ease', a somewhat disturbing result as 'Quality' did not appear to matter to students.

Table 3: Explaining Enrollment Patterns
Y: EnrollOverAve

	(1)	*(2)*	*(3)*
Constant	-2.79 (-3.90)*	-3.67 (-4.01)*	-3.59 (-4.02)*
Time	17.04 (5.59)*	18.50 (4.77)*	18.23 (4.82)*
Time Squared	-17.90 (-5.72)*	-18.66 (-4.70)*	-18.375 (-4.75)*
Evening	.773 (3.56)*	.832 (2.77)*	.817 (2.77)*
Friday	-.175 (-2.81)*	-.095 (-1.05)	-.082 (-.99)
Num. Ratings	.0032 (1.20)	.00046 (.12)	-
Female	-	.071 (.82)	.058 (.72)
Ease	-	.099 (1.62)	.115 (2.55)*
Hot	-	-.080 (-.70)	-
Quality	-	.022 (.41)	-

Table 3: Explaining Enrollment Patterns Y: EnrollOverAve			
	(1)	*(2)*	*(3)*
N	306	145	145
Adj R^2	.111	.137	.153
F	8.59*	3.55*	5.34*
*significant at 5%			

CONCLUDING REMARKS

Research done using data drawn from a single, Mid-Western University confirms that there are decided time preferences in enrollment patterns among multi-sectioned courses where time preferences can be expressed. Further, those time preferences are non-linear with peak demand being in the late morning/early afternoon, with additional demand seen in for evening sections. A slight preference for sections that do not meet on Fridays is also demonstrated.

Faculty members, when identifiable, also affect enrollment patterns. There is no identifiable effect from the gender of the teacher, contrary to conclusions drawn by Centra and Gaubatz, 2000. As for teacher ratings found on the internet, whether they are in themselves valid or not, they do appear to affect enrollment patterns to some extent. Students do not appear to seek out instructors of high quality (either as rated by clarity or helpfulness or the overall combined measure of 'Quality') but do seem to shop for those instructors they believe their peers have rated as 'Easy'. Bringing us back to the old maxim: the grade is all that matters.

REFERENCES

Becker, William E. "Teaching Economics in the 21st Century." *Journal of Economic Perspectives*, 14(1), 2000. pp. 109-119.

Centra, John A. and Noreen B. Gaubatz. "Is there Gender Bias in Student Evaluation of Teaching." *Journal of Higher Education,* 71(1), 2000, 17-33.

Ferber, Marianne A. and Joan A. Huber. "Sex of Student and Instructor: A Study of Student Bias." *American Journal of Sociology* 80(4), 1975, 949-963.

Linsky, Arnold S. and Murray A. Straus. "Student Evaluations, Research Productivity and Eminence of College Faculty." *Journal of Higher Education*, 46(1), 1975, 89-102.

McKeachie, Wilbert S. "Research on College Teaching: The Historical Background." *Journal of Educational Psychology*, 82(2), 1990, 189-200.

Morrow, J. "Some Statistics Regarding the Reliability and Validity of Student Ratings of Teaching." *Research Quarterly,* 48, 1977, 372-75.

Root, Alfred R. "Student Ratings of Teachers." *Journal of Higher Education,* 2(6), 1931, 311-315.

Siegfried, John J. and Peter E. Kennedy. "Does Pedagogy Vary with Class Size in Introductory Economics?" *The American Economic Review*, 85(2), 1995, 347-351.

Taylor, W.S., Frank H. Hankins & Alice A. Lazerowitz. "Student's Ratings of Instructors." *Journal of Higher Education,* 18(4), 1947, 202-206.

ATTITUDES TOWARD TAX EVASION: A DEMOGRAPHIC STUDY OF SOUTH AFRICAN ATTITUDES ON TAX EVASION

Adriana M. Ross, Florida International University
Robert W. McGee, Fayetteville State University

ABSTRACT

A number of studies have examined the relationship between tax collection and various demographic variables. However, until recently most of those studies have involved a United States sample population. The Internal Revenue Service provides demographic data for researchers on a regular basis. The present study goes beyond those studies in several important ways. For one, it uses data on South Africa taken from the World Values database. Not much work has been done on the South African tax or public finance system. Thus, the present study expands on the very limited research done on South African public finance.

The present study expands on existing literature in at least two other ways as well. For one, it examines how various demographics interact with attitudes toward tax evasion. Secondly, we examine several demographic variables that were not examined in prior studies.

One of the questions in the World Values database asked whether it would be justifiable to cheat on taxes if it were possible to do so. Respondents were asked to choose a number from 1 to 10 to indicate the extent of their support for tax evasion. This study examines those responses, both overall and through the prism of more than 20 demographic variables. A trend analysis is also done to determine whether South African attitudes regarding tax evasion have changed in recent years. A comparison is made with other ethical issues to determine the relative seriousness of tax evasion.

The study found that attitudes toward the justifiability of tax evasion often do vary by demographic variable. Tax evasion was found to be a less serious offense than accepting a bribe, suicide or wife beating, equally as serious as prostitution, and more serious than receiving government benefits to which you are not entitled, avoiding a fare on public transport or euthanasia. The trend of opinion on the justifiability of tax evasion has been nonlinear. It is more acceptable in the most recent survey than it was in 1996 but less acceptable than it was in 1990 or 2001.

Although the present study focuses on South Africa, the methodology used in the present study could serve as a template for research on other countries or regions.

INTRODUCTION

Tax evasion has been in existence ever since the first person with sufficient authority attempted to extract tribute from some local population (Adams, 1982, 1993; Webber & Wildavsky, 1986). The topic of tax evasion has become multidisciplinary over the decades, expanding from economics and public finance journals into the fields of accounting, taxation, law, sociology, psychology and other behavioral sciences.

The present study focuses on attitudes toward tax evasion in South Africa. Not many studies have been made of South African public finance and even fewer of attitudes toward tax evasion in South Africa. One study that was made of South African attitudes toward tax evasion was a survey of 191 South African university business students (McGee & Goldman, 2010). That survey included a survey instrument that listed 18 arguments that had been used to justify tax evasion through history. Students were asked to choose a number from one to seven to indicate the extent of their agreement or disagreement with each statement.

The strongest argument to justify tax evasion was in cases where a large portion of the money collected was wasted. The second strongest argument was in cases where the tax system was perceived to be unfair. The next two strongest arguments had to do with oppressive governments – where the taxpayer lived in a repressive regime such as Nazi Germany or Stalinist Russia, or where the government discriminates because of race, religion or ethnic background.

Other strong arguments justifying tax evasion were in cases where the taxpayer was unable to pay, where the government imprisons people for their political opinions, where a significant portion of the money collected winds up in the pockets of corrupt politicians, their friends and families, where the proceeds go to support an unjust war, where tax rates are too high or where funds are spent on projects that the taxpayer morally disapproves of. That study also found that women were more opposed to tax evasion than men, in some cases significantly.

That study also examined the relationship between age and attitude toward tax evasion. The findings were curious, in the sense that they were different from what was expected. Participants in the younger age group (under 25) were more opposed to tax evasion in 10 of 18 cases, whereas the older group (25-40) was more opposed in only 8 of 18 cases. However, in three of those 18 cases the older group was significantly more opposed to tax evasion. The difference in mean scores in the other 15 cases was not significant.

The age findings were curious because they run somewhat contra to the findings in other cases where age was examined. In most other studies that examined age in conjunction with various ethical issues, the older groups were more respectful of the law than were younger groups (Barnett & Karson, 1987, 1989; Longenecker et al., 1989; Harris, 1990; Kelley et al., 1990; Ruegger & King, 1992). Yet in this 2010 South African study, the younger group was more strongly opposed to tax evasion than was the older group in a majority of cases. However, the differences in mean scores were not significant in the cases where the younger group was more opposed, but the scores were significant in three cases where the older group was more

opposed. Part of this lack of tidiness regarding the results could be attributed to the fact that the older group was not that much younger than the younger group.

The prior South African study also examined ethnicity and found that whites were more strongly opposed to tax evasion than Africans in 17 of 18 cases. Mean scores were significantly different in four cases. Other ethnic groups were not included in the statistical calculations due to small sample sizes.

The 2010 South African study also examined religion. The only two religions that had a sufficiently large sample size to compare were Catholics and Other Christians. Catholics were more opposed to tax evasion in 10 of 18 cases. Other Christians were more opposed in 7 cases. In one case, both groups had the same mean scores. None of the differences were significant. The authors of that study speculated that some of the mean score differences might have been significant if the sample size had been larger.

The 2010 study also examined student status as a variable. There were three categories – diploma students, undergraduate and postgraduate students. The diploma students were less opposed to tax evasion than were the other two groups. However, only one difference in mean score was significant – the ability to pay argument. In that case, undergraduate students were more opposed to tax evasion than were the other two groups.

The 2010 study also examined academic major. Management majors were more opposed to tax evasion than economics and finance majors in 11 of 18 cases. However, none of the differences in mean scores were significant. Other business majors were excluded from the mean score comparison because of small sample size.

One good thing that can be said about student surveys, from the researcher's perspective, is that the data is relatively easy to gather. One limiting factor is that the results of student surveys may not be applicable to a wider population. Students are generally younger and more educated than the general population. However, student surveys are a legitimate and popular research methodology and such surveys do make positive contributions to the literatures of many disciplines.

The present study is more than a mere replication of the 2010 South African study. The present study uses the *World Values* survey data that was gathered in South Africa. The sample size is much larger, nearly 3,000. The demographics are also much broader and include a wide range of age groups, races, cultures, occupations, marital status, education and social levels. Because of its comprehensive nature, the present study breaks new ground in several ways. For one, it is more comprehensive than other studies of the South African population. It also examines more demographic variables than any other South African study and also includes more variables than studies of other countries. The methodology used in the present study can be a template for studies of attitudes toward tax evasion in other countries or regions.

REVIEW OF THE LITERATURE

Literature on taxation goes back to the Hebrew Torah and the Christian Bible, at least. In *Genesis* (48: 13-26) there is a description of the Egyptian tax system, which requires farmers to give the Pharaoh 20 percent of their crops. In *Samuel* (I, 8: 15) the tax rate is 10 percent of seed and vineyards. A few lines later (1 *Samuel*, 8:17) it states that the 10 percent rate also extends to sheep and says that taxpayers should consider themselves to be servants.

In *Matthew* (22:21), Jesus says that we should render unto Caesar the things that are Caesar's and to God the things that are God's, but he does not get more specific, leaving future generations without clear guidance as to what the state is entitled to receive in taxes. As a result, a rich body of literature has evolved over the centuries to ponder just what duty individuals have to pay taxes and, perhaps more interestingly, when there is no duty to pay.

Several articles have been written from the Jewish perspective that address the ethics of tax evasion (Cohn, 1998; McGee, 1998g, 1999a; Tamari, 1998; McGee & Cohn, 2008). The Jewish literature is basically strongly against tax evasion. Several reasons are given – "the law is the law;" God commands us to pay taxes; there is a duty not to disparage other Jews and evading taxes makes all Jews look bad; and Jews have a duty to perform good works (mitzvos), which they cannot do if they are in jail, which might be the case if they evade taxes.

McGee (1998a&f, 1999a, 2004, 2012) has criticized all of these rationales. For example, "The law is the law" is not really an argument. Such a statement has an underlying premise that we must obey all laws just because they are laws. Martin Luther King, Gandhi and other rebels, including Jesus, have disputed the view that all laws should automatically and unquestionably be obeyed no matter how stupid or unjust they are. McGee has asked the question several times, "Would it be unethical for Jews living in Nazi Germany to evade taxes if Hitler were the tax collector?" (McGee, 2004, 2006a&b, 2012; McGee & Bernal, 2006; McGee, Nickerson & Fees, 2006, 2009; McGee & López, 2007, 2008; McGee, Alver & Alver, 2008; McGee & Cohn, 2008; McGee & Lingle, 2008; McGee & M'Zali, 2008, 2009; McGee & Preobragenskaya, 2008; McGee & Rossi, 2008; McGee & Tusan, 2008). One might think that such a case would be a deal breaker. One might argue that, surely, in such a case tax evasion would be ethical.

However, when this question was asked in surveys, including the surveys cited in the prior paragraph, respondents generally held that there is some duty for Jews to pay taxes even to Hitler. A survey of Orthodox Jewish students (McGee & Cohn, 2008) found that even Orthodox Jews believe there is some duty to pay taxes to Hitler, although that duty is not absolute.

The Jewish literature on tax evasion, although strongly against tax evasion in general, is not totally opposed to tax evasion (Tamari, 1998). The duty to pay taxes is less than absolute where the ruler is corrupt or where tax funds are squandered.

The literature of other religions has also addressed the tax evasion issue. The Mormon literature is absolutely against tax evasion in all cases. Its literature does not mention a single case where tax evasion can be justified (Smith & Kimball, 1998). However, a survey of Mormon

students found that not all Mormons believe tax evasion can never be justified, although opposition to tax evasion was very strong (McGee & Smith, 2009).

The literature of the Baha'i faith is also strongly against tax evasion (DeMoville, 1998). The only time evasion is permissible is in cases where members of the Baha'i faith are being persecuted.

The Christian literature has also examined the issue of tax evasion. Martin Crowe (1944) wrote a doctoral dissertation that examined 500 years of Catholic literature on the subject. Many of the arguments to justify tax evasion have been discussed in the Catholic literature (McGee, 1998b&c). The ability to pay argument and financing unjust wars are two such arguments, although they are not limited to Catholic scholars.

McGee (1999b) discussed the issue of paying your fair share of taxes. Gronbacher (1998) addressed the issue from the perspective of Catholic social thought and classical liberalism. Pennock (1998) discussed the unjust war excuse for resisting taxes. Schansberg (1998) discussed tax evasion from the perspective of Biblical Christianity, including the issue of what we are obligated to render to Caesar. Some Catholic scholars have said that tax evasion is a mortal sin while others have said it is not. Other Catholic scholars take the position that tax evasion is not a sin at all (Crowe, 1944).

The Muslim literature is mixed on the issue. Two Muslim scholars who wrote books on Islamic business ethics (Ahmad, 1993) and economic justice in Islam (Yusuf, 1971) took the same basic position, that tax evasion is justified if the tax is on income or if the tax causes prices to rise, which would include sales taxes, use taxes and tariffs. They would also justify evasion of death taxes. McGee (1997, 1998d&e, 1999a) mentioned and discussed their views on tax evasion, which caused Jalili (2012) to challenge that position and offer an alternative view of the Muslim literature. According to Jalili, there is an absolute duty to pay taxes to an Islamic state without question, including income taxes, sales and use taxes and tariffs. However, the duty to pay taxes to states that are not purely Islamic, or that are not Islamic at all, is less than absolute. Murtuza and Ghazanfar (1998) do not address those issues, but discuss zakat, the Muslim duty to contribute to the poor.

A number of studies have examined tax evasion from a secular perspective. Armstrong and Robison (1998), McGee (1998f) and Oliva (1998) look at tax evasion from a practitioner perspective. Leiker (1998) discussed Rousseau's views on the issue. The ethics of evading the estate tax (McGee, 1999c) and tariffs (McGee, 1999d) have also been discussed. Ballas and Tsoukas (1998) discuss the reasons for tax evasion in Greece. Other authors have examined reasons for tax evasion in Bulgaria (Smatrakalev, 1998), Russia (Alm, Martinez-Vazquez & Torgler, 2005, 2006; Preobragenskaya & McGee, 2004; Vaguine, 1998) and Armenia (McGee, 1999e). Morales (1998) discussed the case of Mexican street vendors and their view that feeding their families sometimes took precedence over paying taxes.

A few philosophical studies have been done (McGee, 1994, 2004, 2006a, 2012; Thorndike & Ventry, 2002). Nozick (1974) likened the income tax to a form of slavery, since it confiscates the fruits of one's labor.

Others have examined the psychological aspects behind tax evasion (Alm, Martinez-Vazquez & Torgler, 2010; Baird, 1980; Cullis, 2006; Cullis, Jones & Lewis, 2010; Kirchler, 2007). Some authors have written books aimed at the general public (Cowell, 1990; Johnston, 2003, 2007) or the scholarly market (Beito, 1989).

Block conducted two studies that examined the public finance literature in unsuccessful attempts to find a justification for taxation (Block, 1989, 1993). Curry (1982) and Sabrin (1995) discussed the options for financing government without resorting to coercion.

A number of empirical studies have examined tax evasion. Alm and Torgler (2004) examined the determinants of tax morale. They also looked at cultural difference in the USA and Europe (Alm & Torgler, 2006). Cummings, Martinez-Vazquez, McKee and Torgler (2004) also examined the effects of culture on tax compliance, as did Torgler, 2003c). Alm and Martinez-Vazquez (2010) examined tax evasion in the informal sector. Bird, Martinez-Vazquez and Torgler (2004) studied the roe of demand factors in tax performance in developing countries.

Torgler wrote a doctoral dissertation (Torgler 2003a) and a book (Torgler, 2007) on tax morale that included some empirical analyses. He also examined tax morale in transition economies (Torgler, 2003b). Benk, McGee and Ross (2009) did an empirical study of tax evasion opinion in Turkey. Student surveys of tax evasion have also been completed for Argentina (McGee & Rossi, 2008), Armenia (McGee & Maranjyan, 2006), Australia (McGee & Bose, 2009), Bosnia (McGee, Basic & Tyler, 2009), China (McGee & Guo, 2007; McGee & An, 2008), Colombia (McGee, López & Yepes, 2009), Estonia (McGee, Alver & Alver, 2008), France (McGee & M'Zali, 2009), Germany (McGee, Nickerson & Fees, 2006, 2009; McGee, Benk, Ross & Kiliçaslan, 2009),), Guatemala (McGee & Lingle, 2008), Hong Kong (McGee & Butt, 2008; McGee, Ho & Li, 2008), Kazakhstan (McGee & Preobragenskaya, 2008), Macau (McGee, Noronha & Tyler, 2007), Mali (McGee & M'Zali, 2008), New Zealand (Gupta & McGee, 2010), Poland (McGee & Bernal, 2006), Puerto Rico (McGee & López, 2007), Romania (McGee, 2006c), Slovakia (McGee & Tusan, 2008), South Africa (McGee & Goldman, 2010), Taiwan (McGee & Andres, 2009), Turkey (McGee & Benk, 2011), Ukraine (Nasadyuk & McGee, 2008). The limiting factor of student surveys is that the findings may not be generalizable to the general population.

THE PRESENT STUDY

The present study overcomes the limitations of student surveys by including data from a more diverse demographic. It also has a larger sample population than student surveys and examines more demographic variables than is possible with student surveys.

METHODOLOGY

Groups of social scientists all over the world have been conducting coordinated surveys of the world's population since the 1980s. Some surveys have solicited the opinions of more than 200,000 people in more than 80 countries. The surveys included hundreds of questions on a wide range of subjects. One question in the most recent surveys addressed attitudes toward tax evasion:

Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something in between: Cheating on taxes if you have a chance.

The range of responses used a 10-point Likert Scale where 1 = never justifiable and 10 = always justifiable. The surveys collected data on a number of demographic variables, including level of education, gender and age. The present study uses the data gathered in the most recent survey on South Africa. The sample size was nearly 3,000.

More that 20 demographic variables are examined using t-tests and ANOVAs to determine whether any differences are significant at the 5 percent level. The ANOVA was used to analyze mean score differences between groups as a whole. The ANOVA scores are reported in the "b" tables. T-tests were sometimes made to compare the mean scores of two particular groups. Those scores, where made, are reported in the "a" tables.

FINDINGS

The findings are reported below for more than 20 demographic variables.

Gender

Table 1 shows the results for gender. Women were significantly more opposed to tax evasion ($p = 0.0181$). This finding is similar to the findings in studies of Australia (McGee & Bose, 2009), China (McGee & Guo, 2007), Colombia (McGee, López & Yepes, 2009), Estonia (McGee, Alver & Alver, 2008), Germany (McGee, Nickerson & Fees, 2006), Guatemala (McGee & Lingle, 2008), Orthodox Jewish students (McGee & Cohn, 2008), international business academics (McGee, 2006b), New Zealand (Gupta & McGee, 2010), Puerto Rico (McGee & López, 2007), Taiwan (McGee & Andres, 2009) and Thailand (McGee, 2008).

However, not all studies on tax evasion that have examined gender differences found that women are more opposed to tax evasion. Men were found to be more opposed to tax evasion in Romania (McGee, 2006c), Slovakia (McGee & Tusan, 2008), Turkey (Benk, McGee & Ross, 2009; McGee & Benk, 2011McGee, Benk, Yildirim & Kayikçi, 2011) and Vietnam (McGee, 2008a). Other studies found no statistical difference between male and female opinions on tax evasion – Argentina (McGee & Rossi, 2008), China (McGee & An, 2008; McGee & Noronha,

2008), France (McGee & M'Zali, 2009), Hong Kong (McGee & Butt, 2008), Kazakhstan (McGee & Preobragenskaya, 2008), Macau (McGee, Noronha & Tyler, 2007).

H1: People are equally averse to tax evasion regardless of gender.
H1: Rejected.

Table 1: Ranking by Gender
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Gender	Mean	Std. Dev.	n
1	Female	2.4	2.23	1456
2	Male	2.6	2.33	1455
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Male v. Female		0.0181		

Age

Tables 2a and 2b show the results for the age demographic. The three oldest age groups had the three lowest mean scores, meaning they had the strongest opposition to tax evasion. The three youngest groups had the least opposition to tax evasion. However, the ANOVA found that the differences in mean scores between groups was not significant ($p = 0.352$). Some of the t-tests of individual comparisons found significant differences at the 10 percent level ($p = 0.0614$ & 0.0841).

This finding runs contra to the findings of some other studies. Some studies that analyzed ethics and age found that older people are more ethical than younger people or that older people have more respect for the law and for authority than do young people (Babakus et al., 2004; Barnett & Karson, 1987, 1989; Longenecker wt al., 1989; Harris, 1990; Kelley et al., 1990; Ruegger & King, 1992; Serwinek, 1992, Wood et al., 1988).

However, not all studies found that the relationship between age and ethics was positive. Some studies found that younger people are more ethical (Babakus et al., 2004; Browning & Zabriskie, 1983; Sims et al., 1996). Other studies found no statistical correlation between age and ethical beliefs (Kidwell et al., 1987; Izraeli, 1988; Callan, 1992; Kohut & Corriher, 1994).

H2: People are equally averse to tax evasion regardless of age.
H2: Cannot be rejected.

Table 2a: Ranking by Age (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Age**	**Mean**	**Std. Dev.**	**n**
1	55-64	2.3	2.16	263
2	45-54	2.4	2.14	331
2	65+	2.4	2.31	235
4	15-24	2.5	2.31	790
5	25-34	2.6	2.24	708
5	35-44	2.6	2.41	583
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
25-34 v. E55-64		0.0614		
35-44 v. E55-64		0.0841		

Table 2b: Age and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	28.954	5	5.791	1.112	0.352
Within Groups	15,120.243	2,904	5.207		
Total	15,149.197	2,909			

Marital Status

The data on marital status is shown in Tables 3a and 3b. The group most opposed to tax evasion was divorced people. In second place was married people, followed by widowed people. The group least opposed to tax evasion was people who were living together as married. The difference in mean scores between groups was highly significant ($p < 0.0001$).

A comparative study of tax evasion in Moldova and Romania found that married people were significantly more opposed to tax evasion than were single people in Moldova; married and divorced people were equally opposed to tax evasion; divorced and single people were equally opposed to tax evasion (McGee, 2009). The Romanian sample found that married people were significantly more opposed to tax evasion than were single people; there was no significant difference between married v. divorced or divorced v. single people.

H3: People are equally averse to tax evasion regardless of marital status.
H3: Rejected.

Table 3a: Ranking by Marital Status (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Marital Status**	**Mean**	**Std. Dev.**	**n**
1	Divorced	1.9	1.94	61
2	Married	2.3	2.09	1021
3	Widowed	2.4	2.12	177
4	Separated	2.5	2.26	32
5	Single/Never married	2.6	2.39	1311
6	Living together as married	2.9	2.50	309
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Married v. Living together as married		0.0001		
Married v. Single/Never married		0.0015		
Living together as married v. Divorced		0.0034		
Living together as married v. Widowed		0.0256		
Living together as married v. Single/Never married		0.0493		
Divorced v. Single/Never married		0.0244		

Table 3b: Marital Status and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	127.115	5	25.423	4.911	<0.0001
Within Groups	15,038.479	2,905	5.177		
Total	15,165.594	2,910			

Number of Children

Tables 4a and 4b show the data for number of children. The group most opposed to tax evasion was the group that had 6 children. Those who had 8 or more children ranked second. Those who had 0, 1 or 3 children were least opposed to tax evasion. There seemed to be no clear pattern, in the sense that the relationship was not linear. An ANOVA found that the difference in mean score between groups was highly significant (p = 0.006).

H4: People are equally averse to tax evasion regardless of number of children.
H4: Rejected.

Table 4a: Ranking by Number of Children (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Number of Children**	**Mean**	**Std. Dev.**	**n**
1	6	1.5	1.31	59
2	8 or more	2.0	1.69	36
3	5	2.2	1.81	127
4	2	2.4	2.11	614
4	7	2.4	2.00	32
6	4	2.5	2.35	206
7	None	2.6	2.38	923
7	1	2.6	2.33	515
7	3	2.6	2.46	401
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
None v. 6		0.0005		
1 v. 6		0.0004		
2 v. 6		0.0014		
3 v. 6		0.0008		
4 v. 6		0.0019		
5 v. 6		0.0084		
6 v. 7		0.0113		

Table 4b: Number of Children and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	103.941	7	14.849	2.862	0.006
Within Groups	14,907.207	2,873	5.189		
Total	15,011.148	2,880			

Religion

Tables 5a and 5b show the results for religion. Although Jews were ranked as being most opposed to tax evasion, the small sample size makes the result unreliable. Of groups where the sample size was 30 or more, Muslims ranked first, followed by Pentecostals and Evangelicals. Jehovah witnesses were least opposed to tax evasion. An ANOVA showed that the mean difference between groups was significant at the 10 percent level ($p = 0.063$). A comparison

between two individual groups using t-tests sometimes found differences at the 1 percent or 5 percent level, as is seen at the bottom of Table 5a. Other comparisons were significant only at the 10 percent level.

H5: People are equally averse to tax evasion regardless of religion.

H5: Rejected. Although the ANOVA showed no difference at the 5 percent level, some individual comparisons showed significant difference at the 1 percent and 5 percent levels.

Table 5a: Ranking by Religion
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Religion	Mean	Std. Dev.	n
1	Jew	1.3	0.93	3
2	Muslim	1.7	1.58	36
3	Buddhist	1.8	1.02	4
4	Orthodox	2.1	1.98	6
5	Pentecostal	2.3	2.13	61
6	Evangelical	2.4	2.24	282
6	Independent African Church	2.4	2.14	490
6	Other	2.4	2.15	193
6	Protestant	2.4	2.31	865
10	Roman Catholic	2.7	2.26	356
11	Hindu	2.8	2.39	29
12	Jehovah	3.1	2.73	45

SIGNIFICANT DIFFERENCES IN MEAN SCORES

		p value	Significant?
Hindu v. Muslim		0.0297	Yes, at 5%
Jehovah v. Muslim		0.0078	Yes, at 1%
Jehovah v. Protestant		0.0499	Yes, at 5%
Muslim v. Roman Catholic		0.0100	Yes, at 1%
Protestant v. Roman Catholic		0.0382	Yes, at 5%
Evangelical v. Jehovah		0.0602	Yes, at 10%
Evangelical v. Muslim		0.0702	Yes, at 10%
Evangelical v. Roman Catholic		0.0951	Yes, at 10%
Jehovah v. Pentecostal		0.0931	Yes, at 10%
Muslim v. Protestant		0.0722	Yes, at 10%

Table 5b: Religion and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	67.422	7	9.632	1.918	0.063
Within Groups	11,647.994	2,320	5.021		
Total	11,715.416	2,327			

Religious Practice

This question asked how often do you attend religious services? Tables 6a and 6b show the results. The results were interesting and did not follow a clear pattern. The two groups that were most opposed to tax evasion were the groups that attended religious services once a year and more than once a week, which were at opposite ends of the spectrum. Those who attended religious services once a week and once a month had similar mean scores. However, the groups ranked in the three last places had the least opposition to tax evasion. The ANOVA found that the mean score differences between groups was significant. ($p < 0.0001$).

H6: People are equally averse to tax evasion regardless of religious practice.
H6: Rejected.

Table 6a: Ranking by Religious Practice (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Religious Practice**	**Mean**	**Std. Dev.**	**n**
1	Once a year	1.7	1.58	32
2	More than once a week	2.1	2.04	488
3	Once a month	2.4	2.14	433
4	Once a week	2.5	2.27	1093
5	Never/practically never	2.7	2.52	395
6	Less than once a year	2.9	2.27	296
7	Only on special holy days	3.1	2.64	176
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value	**Sign?**	
More than once a week v. Once a week		0.0009	1%	
More than once a week v. Once a month		0.0298	5%	
More than once a week v. Only on special holy days		0.0001	1%	
More than once a week v. Less than once a year		0.0001	1%	
More than once a week v. Never/practically never		0.0001	1%	
Once a week v. Only on special holy days		0.0015	1%	

Table 6a: Ranking by Religious Practice (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)			
Once a week v. Once a year	0.0480	5%	
Once a week v. Less than once a year	0.0072	1%	
Once a month v. Only on special holy days	0.0007	1%	
Once a month v. Less than once a year	0.0026	1%	
Only on special holy days v. Once a year	0.0041	1%	
Once a year v. Less than once a year	0.0038	1%	
Once a year v. Never/practically never	0.0277	5%	
Once a month v. Once a year	0.0704	10%	
Once a month v. Never/practically never	0.0645	10%	
Only on special holy days v. Never/practically never	0.0849	10%	

Table 6b: Religious Practice and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	228.891	6	38.148	7.415	<0.0001
Within Groups	14,951.285	2,906	5,145		
Total	15,180.175	2,912			

Importance of God in Your Life

The question asked "How important is God in your life?" Table 7a and 7b show the data. The ANOVA found that the difference between groups was highly significant ($p < 0.0001$). However, it was difficult to see a clear-cut pattern. The two groups most opposed to tax evasion were groups 2 and 10, which are at opposite ends of the spectrum.

H7: People are equally averse to tax evasion regardless of the importance of God in their life.
H7: Rejected.

Table 7a: Ranking by Importance of God in Your Life (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Importance of God in Your Life**	**Mean**	**Std. Dev.**	**n**
1	2	2.2	1.09	7
2	10 Very important	2.3	2.23	1896
3	3	2.5	2.59	11
3	5	2.5	2.09	85

Table 7a: Ranking by Importance of God in Your Life (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Importance of God in Your Life**	**Mean**	**Std. Dev.**	**n**
3	9	2.5	2.12	309
6	8	2.9	2.36	307
7	6	3.1	2.18	78
8	7	3.4	2.62	161
9	1 Not at all important	3.5	3.25	13
10	4	4.9	2.32	24
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
1 Not at all important v. 10 Very important		0.0542		
4 v. 5		0.0001		
4 v. 6		0.0007		
4 v. 7		0.0087		
5 v. 7		0.0066		
6 v. 7		0.0019		
6 v. 9		0.0269		
6 v. 10 Very important		0.0019		
5 v. 6		0.0748		

Table 7b: Importance of God in Your Life and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	434.678	7	62.097	12.191	<0.0001
Within Groups	14,593.923	2,865	5.094		
Total	15,028.601	2,872			

Education Level

Tables 8a and 8b show the data for education level. Those with university degrees were most opposed to tax evasion. Those who had some university education were least opposed. Thus, the two most educated groups were at opposite ends of the scale. Other than that, there seemed to be a pattern that the most educated groups were most opposed to tax evasion and those with the least education were least opposed to tax evasion. The ANOVA was highly significant ($p < 0.0001$).

H8: People are equally averse to tax evasion regardless of education level.
H8: Rejected.

Table 8a: Ranking by Education Level (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Education Level**	**Mean**	**Std. Dev.**	**n**
1	University with degree	1.9	1.79	132
2	Complete secondary – technical, vocational	2.0	1.95	198
3	Complete secondary – college preparatory	2.4	2.16	646
4	Incomplete secondary – college preparatory	2.5	2.24	1070
5	No formal education	2.6	2.14	224
6	Completed elementary	2.8	2.46	241
7	Inadequately completed elementary education	2.9	2.72	356
8	Incomplete secondary – technical, vocational	3.1	2.51	25
9	Some university without degree	3.3	2.68	16
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
			p value	
No formal education v. Complete secondary – technical, vocational			0.0029	
No formal education v. University with degree			0.0017	
Completed elementary v. Complete secondary – technical, vocational			0.0002	
Completed elementary v. Complete secondary – college preparatory			0.0185	
Completed elementary v. University with degree			0.0002	
Complete secondary – technical, vocational v. Complete secondary – college preparatory			0.0200	
Complete secondary – college preparatory v. University with degree			0.0130	

Table 8b: Education Level and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	193.340	7	27.620	5.385	<0.0001
Within Groups	14,793.225	2,884	5.129		
Total	14,986.565	2,891			

Employment Status

Tables 9a and 9b show the data for the employment status category. Housewives was the group most opposed to tax evasion. There was a three-way tie for second place – self-employed,

retired and students. Part-timers and unemployed were the two groups who were least opposed to tax evasion. The ANOVA found the differences to be highly significant ($p = 0.0004$). T-test comparisons found that some comparisons between individual groups were significant at the 1 percent and 5 percent levels, while other comparisons were significant at the 10 percent level.

One may only speculate about the reasons for this ranking. Perhaps housewives are strongly opposed to tax evasion because of gender influences. Women often tend to be more strongly opposed to tax evasion than men and more respectful of authority in general. Retired people might be more firmly opposed to tax evasion than the general population because of their age. Older people tend to be more opposed to tax evasion and more respectful of the law and authority. What is somewhat surprising is that self-employed people are so firmly against tax evasion. They have more opportunities to evade taxes and one might think that they would be less opposed to tax evasion, but such was not the case. One might also think that students would be less opposed to tax evasion, since they are young, and some studies have shown that younger people are generally less respectful of authority and less ethical than are older people.

H9: People are equally averse to tax evasion regardless of employment status.
H9: Rejected.

Table 9a: Ranking by Employment Status
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Employment Status	Mean	Std. Dev.	n
1	Housewife	2.0	1.85	152
2	Self employed	2.4	2.14	116
2	Retired	2.4	2.33	338
2	Students	2.4	2.31	400
5	Full time	2.5	2.23	787
6	Unemployed	2.7	2.29	894
7	Part time	2.8	2.56	223
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Full time v. Housewife		0.0096		
Part time v. Housewife		0.0010		
Part time v. Students		0.0468		
Retired v. Unemployed		0.0414		
Housewife v. Unemployed		0.0004		
Students v. Unemployed		0.0300		
Full time v. Part time		0.0868		
Full time v. Unemployed		0.0707		

Table 9a: Ranking by Employment Status (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)			
Part time v. Retired	0.0563		
Retired v. Housewife	0.0624		
Housewife v. Students	0.0561		

Table 9b: Employment Status and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	99.928	6	16.655	3.213	0.004
Within Groups	15,048.675	2,903	5.184		
Total	15,148.603	2,909			

Occupation

Tables 10a and 10b show the data for the occupation demographic. The two groups most opposed to tax evasion were supervisory non-manual office workers and non-manual office workers. Office workers and professionals tended to be more averse to tax evasion than other groups, in general. Groups least opposed to tax evasion were farmers and agricultural workers and various categories of manual workers. If one could generalize, one might say that people who work with their hands are less averse to tax evasion than people who work with their brains. The ANOVA found the differences between groups to be highly significant ($p < 0.0001$). Many of the t-test of individual group comparisons also found a high level of significance.

H10: People are equally averse to tax evasion regardless of occupation.
H10: Rejected.

Table 10a: Ranking by Occupation (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Occupation**	**Mean**	**Std. Dev.**	**n**
1	Supervisory non-manual office worker	1.7	1.38	75
1	Non-manual office worker	1.7	1.41	157
3	Employer/manager of establishment with 10 or more employed	1.8	2.07	57
3	Professional worker	1.8	1.64	144
3	Never had a job	1.8	1.28	32
6	Employer/manager of establishment with less than 10 employed	2.0	1.86	42

Table 10a: Ranking by Occupation (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Occupation**	**Mean**	**Std. Dev.**	**n**
7	Member of armed forces	2.4	2.32	76
8	Foreman and supervisor	2.5	2.42	35
8	Semi-skilled manual worker	2.5	2.33	216
10	Skilled manual	2.8	2.47	214
10	Unskilled manual worker	2.8	2.52	554
12	Farmer – has own farm	3.0	2.39	11
13	Agricultural worker	3.1	2.41	140

SIGNIFICANT DIFFERENCES IN MEAN SCORES	
	p value
Employer/manager of establishment with 10 or more employed v. Skilled manual	0.0054
Employer/manager of establishment with 10 or more employed v. Unskilled manual worker	0.0039
Professional worker v. Foreman and supervisor	0.0423
Professional worker v. Skilled manual	0.0001
Professional worker v. Semi skilled manual worker	0.0019
Professional worker v. Agricultural worker	0.0001
Professional worker v.Member of armed forces	0.0271
Supervisory non-manual office worker v. Foreman and supervisor	0.0298
Supervisory non-manual office worker v. Skilled manual	0.0003
Supervisory non-manual office worker v. Agricultural worker	0.0001
Supervisory non-manual office worker v. Member of armed forces	0.0259
Agricultural worker v. Member of armed forces	0.0401
Agricultural worker v. Never had a job	0.0036

Table 10b: Occupation and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	339.056	7	48.437	9.464	<0.0001
Within Groups	8,025.164	1,568	5.118		
Total	8,364.220	1,575			

Institution of Occupation

Tables 11a and 11b show the results for the institution of occupation category. One might reason a priori that people who work for government institutions would be more averse to tax evasion than the general population, since their salaries are paid out of tax funds. Alternatively, one might guess that government workers would be less averse to tax evasion than the general population, since seeing how government works from the inside can cause one to become cynical.

An analysis of mean scores found that employees of public institutions (government workers) were more strongly opposed to tax evasion than were the other groups, although a t-test of the mean scores comparing public institutions to private business found that the difference in mean scores was not significant. However, the mean score for private non-profit organization was significantly higher than the mean scores of the other two categories. An ANOVA found the between group differences to be significant at the 1 percent level ($p < 0.0001$).

H11: People are equally averse to tax evasion regardless of institution of occupation.
H11: Rejected.

Table 11a: Ranking by Institution Of Occupation
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Institution of Occupation	Mean	Std. Dev.	n
1	Public Institution	2.2	2.18	279
2	Private Business	2.4	2.25	1249
3	Private Non-profit Organization	3.1	2.43	230
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Public Institution v. Private Non-profit Organization		0.0001	Yes	
Private Business v. Private Non-profit Organization		0.0001	Yes	

Table 11b: Institution of Occupation and Attitudes toward Tax Evasion
ANOVA Analysis

	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	117.565	2	58.782	11.474	<0.0001
Within Groups	8,991.389	1,755	5.123		
Total	9,108.954	1,757			

Income Level

Tables 12a and 12b show the data for the income level category. One might assume a priori that the more income people earn, the less averse they are to tax evasion, since they probably perceive that they are not getting their money's worth from government services. However, unless one conducts a test, such a view is mere speculation. Table 12a shows the mean score ranking for the ten categories of income level. The ranking reveals an interesting and unexpected relationship. The middle-income groups tend to be more opposed to tax evasion than either the low-income or high-income groups. One possible explanation of this pattern might be that people at the low-income levels have less respect for authority, or perhaps even antagonism for government, and people at the high-income levels might feel exploited by the graduated income tax, and that the benefits they receive from government are less than the taxes they have to pay to support the system. The ANOVA found the differences between groups to be highly significant ($p < 0.0001$).

H12: People are equally averse to tax evasion regardless of income level.
H12: Rejected.

Table 12a: Ranking by Income Level
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Income Level	Mean	Std. Dev.	n
1	Sixth step	2.1	1.88	359
2	Fifth step	2.3	2.03	405
3	Seventh step	2.4	2.11	291
4	Fourth step	2.5	2.16	365
5	Third step	2.6	2.22	302
6	Eighth step	2.7	2.43	236
7	Lower step	2.8	2.71	401
7	Second step	2.8	2.35	295
9	Ninth step	2.9	2.65	54
10	Tenth step	3.5	3.32	52
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Lower step v. 5		0.0031		
Lower step v. 6		0.0001		
Lower step v. 10		0.0889		
3 v. 10		0.0134		
5 v. 8		0.0258		

Table 12a: Ranking by Income Level (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)			
5 v. 10	0.0003		
7 v. 10	0.0019		
8 v. 10	0.0464		
L v. 4	0.0927		
3 v. 5	0.0623		
5 v. 9	0.0504		

Table 12b: Income Level and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	p value
Between Groups	151.252	7	21.607	4.282	<0.0001
Within Groups	13,351.901	2,646	5.046		
Total	13,503.153	2,653			

Ethnicity

The World Values data placed ethnicity into four categories. Tables 13a and 13b show the data for the ethnicity demographic. Whites were most opposed to tax evasion. South Asians and blacks were least opposed. The ANOVA found that the differences between groups were significant at the 1 percent level ($p < 0.0001$). T-tests of individual groups found that the mean score differences between whites and coloured was statistically insignificant.

H13: People are equally averse to tax evasion regardless of ethnicity.
H13: Rejected.

Table 13a: Ranking by Ethnicity (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Ethnicity**	**Mean**	**Std. Dev.**	**n**
1	White	1.9	1.80	321
2	Coloured (dark)	2.0	2.00	205
3	Asian South	2.6	2.19	59
3	Black	2.6	2.35	2326

Table 13a: Ranking by Ethnicity (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Asian South v. Coloured (dark)		0.0479		
Asian South v. White		0.0084		
Black v. Coloured (dark)		0.0001		
Black v. White		0.0004		

Table 13b: Ethnicity and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	189.559	3	63.186	12.269	<0.0001
Within Groups	14,970.786	2,907	5.510		
Total	15,160.346	2,910			

Social Class

Tables 14a and 14b show the results for the social class demographic. Working class people tended to be the most averse to tax evasion, while people from the lower class tended to be least averse to tax evasion, which is an interesting result, since those two classes seemingly have a lot in common. The upper and middle class groups tended to rank in the middle. The ANOVA found the difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H14: People are equally averse to tax evasion regardless of social class.
H14: Rejected.

Table 14a: Ranking by Social Class (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Social Class**	**Mean**	**Std. Dev.**	**n**
1	Working class	2.1	1.94	528
2	Upper middle class	2.3	2.13	430
3	Upper class	2.4	2.14	68
4	Lower middle class	2.6	2.37	544
5	Lower class	2.8	2.44	1202

Table 14a: Ranking by Social Class (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Upper middle class v. Lower middle class		0.0406		
Upper middle class v. Lower class		0.0002		
Lower middle class v. Working class		0.0002		
Working class v.Lower class		0.0001		

Table 14b: Social Class and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	211.535	4	52.884	10.136	<0.0001
Within Groups	14,436.831	2,767	5.218		
Total	14,648.366	2,771			

Region

Tables 15a and 15b show the data by region. A priori it is difficult to guess what the results might be by region unless one is an expert on the regions of South Africa. People who live in Western Cape were more opposed to tax evasion than were residents of any other region. Those who lived in Northern Cape were least opposed to tax evasion. The ANOVA found the mean score difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H15: People are equally averse to tax evasion regardless of region.
H15: Rejected.

Table 15a: Ranking by Region (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Region**	**Mean**	**Std. Dev.**	**n**
1	Western Cape	1.8	1.60	271
2	Eastern Cape	2.2	2.08	454
3	KwaZulu	2.4	2.25	624
4	Gauteng	2.5	2.22	560
5	Mpumalanga	2.6	2.25	192
6	Northern Province	2.9	2.60	357

Table 15a: Ranking by Region (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Region**	**Mean**	**Std. Dev.**	**n**
6	North West	2.9	2.34	244
8	Free State	3.2	2.40	182
9	Northern Cape	4.5	3.67	26
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Northern Province v. KwaZulu		0.0016		
Northern Province v. Gauteng		0.0131		
Northern Province v. Eastern Cape		0.0001		
Northern Province v. Northern Cape		0.0035		
Northern Province v. Western Cape		0.0001		
KwaZulu v. Northern Cape		0.0001		
KwaZulu v. Western Cape		0.0001		
KwaZulu v. Free State		0.0001		
KwaZulu v. North West		0.0037		
Gauteng v. Northern Cape		0.0001		
Gauteng v. Western Cape		0.0001		
Gauteng v. Free State		0.0003		
Gauteng v. North West		0.0211		
Mpumalanga v. Eastern Cape		0.0297		
Mpumalanga v. Northern Cape		0.0003		
Mpumalanga v. Western Cape		0.0001		
Mpumalanga v. Free State		0.0130		
Eastern Cape v. Northern Cape		0.0001		
Eastern Cape v. Western Cape		0.0067		
Eastern Cape v. Free State		0.0001		
Eastern Cape v. North West		0.0001		
Northern Cape v. Western Cape		0.0001		
Northern Cape v. Free State		0.0175		
Northern Cape v. North West		0.0021		
Western Cape v. Free State		0.0001		
Western Cape v. North West		0.0001		

Table 15b: Region and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	367.149	7	52.450	10.544	<0.0001
Within Groups	14,306.601	2,876	4.974		
Total	14,673.750	2,883			

Feeling of Happiness

Tables 16a and 16b show the data for the happiness categories. One might speculate a priori that the happiest people are the people who evade taxes because they get to keep more of the fruits of their labor. Alternatively, one might speculate that those who do not evade taxes are happiest because they have peace of mind. The ranking showed that people who are very happy and people who are not very happy are equally opposed to tax evasion. Those who are not happy at all are least opposed to tax evasion. However, the ANOVA found that the differences between groups are significant only at the 10 percent level (p = 0.091). A t-test comparison found that the very happy group was significantly more opposed to tax evasion than was the quite happy group (p = 0.0386). However, a comparison of the very happy and not happy at all groups found the difference in mean scores not to be significant, even at the 10 percent level (p = 0.1064).

H16: People are equally averse to tax evasion regardless of level of happiness.
H16: Rejected.

Table 16a: Ranking by Feeling of Happiness (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Feeling of Happiness**	**Mean**	**Std. Dev.**	**n**
1	Very happy	2.4	2.24	1250
1	Not very happy	2.4	2.14	454
3	Quite happy	2.6	2.36	1028
4	Not happy at all	2.7	2.49	170
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Very happy v. Quite happy		0.0386		

Table 16b: Feeling of Happiness and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	33.731	3	11.244	2.157	0.091
Within Groups	15,109.337	2,898	5.214		
Total	15,143.068	2,901			

Satisfaction with Life

Results for the satisfaction with life demographic are shown in Tables 17a and 17b. The ranking does not show a discernible pattern. Those most dissatisfied with life are most opposed to tax evasion, while those who are almost as dissatisfied with life (group 3) are least opposed to tax evasion. The ANOVA found the differences between groups to be significant at the 1 percent level ($p < 0.0001$).

H17: People are equally averse to tax evasion regardless of their degree of satisfaction with life.

H17: Rejected.

Table 17a: Ranking by Satisfaction with Life (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Satisfaction with Life**	**Mean**	**Std. Dev.**	**n**
1	1 Dissatisfied	2.1	2.36	99
2	2	2.2	2.12	75
2	5	2.2	2.06	288
4	10 Satisfied	2.3	2.24	510
5	8	2.4	2.04	528
6	6	2.5	2.28	328
7	7	2.7	2.08	418
7	9	2.7	2.45	404
9	4	3.1	2.71	154
10	3	4.0	2.96	97
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
3 v. 8		0.0001		
5 v. 6		0.0888		
2 v. 9		0.0984		
4 v. 7		0.0617		

Table 17b: Satisfaction with Life and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	155.549	7	22.221	4.460	<0.0001
Within Groups	13,557.496	2,721	4.983		
Total	13,713.045	2,728			

State of Health

Tables 18a and 18b show the results for the health category. Those who were in fair health were most opposed to tax evasion, while those in poor health were least opposed to tax evasion. However, the differences in mean scores were insignificant, both when using the ANOVA (p = 0.563) and when conducting t-test of two individual categories.

H18: People are equally averse to tax evasion regardless of health status.

H18: Cannot be rejected.

Table 18a: Ranking by State of Health (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**State of Health**	**Mean**	**Std. Dev.**	**n**
1	Fair	2.4	2.21	415
2	Very good	2.5	2.34	1179
2	Good	2.5	2.20	1107
4	Poor	2.7	2.47	162
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
None Significant				

Table18b: State of Health and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	10.601	3	3.534	0.682	0.563
Within Groups	14,807.559	2,859	5.179		
Total	14,818.160	2,862			

Self Positioning in Political Scale

Tables 19a and 19b show the data from the political positioning scale. This variable was tested to determine whether those on the left had views that differed from those on the right. One might speculate a priori that those on the left (socialists) would be more opposed to tax evasion than those on the right (capitalists) because of the belief that there is a duty to the state. However, it is mere speculation unless the theory is tested.

The ranking of mean scores shows that the three groups most opposed to tax evasion are all left of center (lower than 6) and the two groups least opposed to tax evasion are right of center (greater than 5), which lends support to the a priori relationship. Some of the centrist groups tend to be in the middle of the ranking. However, the farthest right group (10) is also in the middle, which does not support the a priori theory. The ANOVA found the difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H19: People are equally averse to tax evasion regardless of position on the political scale.
H19: Rejected.

Table 19a: Ranking by Self Positioning in Political Scale
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Self Positioning in Political Scale	Mean	Std. Dev.	n
1	1 Left	2.1	2.25	56
1	3	2.1	1.84	94
3	5	2.2	2.00	358
4	6	2.3	2.06	402
5	4	2.5	2.11	157
5	10 Right	2.5	2.48	304
7	7	2.6	2.16	327
8	2	2.7	2.14	50
9	9	3.0	2.80	241
10	8	3.1	2.66	384
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
1 Left v. 9		0.0253		
1 Left v. 8		0.0076		
3 v. 8		0.0006		
3 v. 7		0.0419		

Table 19b: Self Positioning in Political Scale and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	254.060	7	36.294	6.767	<0.0001
Within Groups	12,115.184	2,259	5.363		
Total	12,369.244	2,266			

Hard Work Brings Success

Tables 20a and 20b show the results for the view that hard work brings success. One might speculate a priori that those who believe that hard work brings success might also be less opposed to tax evasion, since they probably also perceive that they are being unjustly deprived of the fruits of their labor, whereas those who believe that luck and connections are the reasons for success might have a cynical attitude that carries over into the realm of tax evasion. However, such an a priori conclusion is only speculative unless tested. All 5 of the top 5 groups were ranked in the top 6, indicating that these groups are more strongly opposed to tax evasion than is the general population. The ANOVA found the differences between groups to be significant at the 1 percent level ($p < 0.0001$).

H20: People are equally averse to tax evasion regardless of attitude toward hard work.
H20: Rejected.

Table 20a: Ranking by Hard Work Brings Success (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Hard Work Brings Success**	**Mean**	**Std. Dev.**	**n**
1	1 In the long run, hard work usually brings a better life	2.1	2.01	890
2	3	2.3	2.08	306
3	2	2.5	2.28	470
3	4	2.5	2.05	226
3	10 Hard work doesn't generally bring success – \ it's more a matter of luck and connections.	2.5	2.76	121
6	5	2.6	2.12	241
7	6	3.0	2.45	249
7	9	3.0	2.36	77
9	8	3.2	2.45	140
10	7	3.9	3.00	149

Table 20a: Ranking by Hard Work Brings Success (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
3 v. 8		0.0001		
4 v. 7		0.0001		
1 In the long run, hard work usually brings a better life v. 10 Hard work doesn't generally bring success – it's more a matter of luck and connections		0.0510		
2 v. 9		0.0765		
5 v. 6		0.0542		

Table 20b: Hard Work Brings Success and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	579.399	7	82.771	16.918	<0.0001
Within Groups	13,028.437	2,663	4.892		
Total	13,607.836	2,670			

Income Equality

Tables 21a and 21b show the results for the relationship between the belief that incomes should be more equal and attitude toward tax evasion. One may speculate a priori that those who favor more income equality would want to use the tax system to achieve their goal and that they would be more opposed to tax evasion than the general population. The ranking tends to confirm that view. Those in the three most extreme "incomes should be more equal" categories were the three groups most opposed to tax evasion. However, those who were most opposed to the equal incomes position (group 10) ranked fourth, which does not support the a priori view. The ANOVA found that the difference between groups was significant at the 1 percent level ($p < 0.0001$).

H21: People are equally averse to tax evasion regardless of attitude toward equality of income.
H21: Rejected.

Table 21a: Ranking by Income Equality (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Income Equality**	**Mean**	**Std. Dev.**	**n**
1	1 Incomes should be made more equal	2.0	1.97	478
2	3	2.1	1.61	172
3	2	2.3	2.07	261
4	10 We need larger income differences as incentives	2.5	2.56	404
5	7	2.6	2.23	233
6	5	2.7	2.13	283
6	8	2.7	2.32	311
8	4	2.8	2.16	179
8	6	2.8	2.43	270
10	9	3.1	2.86	257
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
1 Incomes should be made more equal v. 10 We need larger income differences as incentives		0.0011		
2 v. 9		0.0003		
3 v. 8		0.0027		

Table 21b: Income Equality and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	269.881	7	38.554	7.166	<0.0001
Within Groups	13,390.422	2,489	5.380		
Total	13,660.304	2,496			

Private vs. State Ownership of Business

Tables 22a and 22b Show the results for the view on ownership of business. One might speculate a priori that those who favor more government ownership of business (socialist position) would be more opposed to tax evasion than those who favor more private ownership (capitalist position), since socialists tend to favor more government ownership, which requires tax funding. The ranking of mean scores tends to reject this a priori theory, since four of the top 5 ranks (2-1-3-4) could be labeled as capitalist. The ANOVA found the difference between

groups to be significant at the 1 percent level ($p < 0.0001$). Thus, those who favor more private ownership of business tend to be more opposed to tax evasion than those who favor more government ownership of business, which is somewhat of a surprising result. One possible explanation for this finding is that those who favor the private sector over the government sector have stronger support for property rights and the rule of law, which includes the belief that one should not evade taxes.

H22: People are equally averse to tax evasion regardless of attitude toward ownership of business.
H22: Rejected.

Table 22a: Ranking by Private V. State Ownership of Business
(Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)

Rank	Private v. State Ownership of Business	Mean	Std. Dev.	n
1	2	1.7	1.35	309
2	1 Private ownership of business should be increased	1.8	1.99	328
3	10 Government ownership of business should be increased	2.2	2.35	420
4	3	2.3	1.88	204
5	4	2.6	2.16	192
6	6	2.9	2.33	304
6	8	2.9	2.30	227
8	5	3.0	2.50	372
9	7	3.1	2.34	245
10	9	3.2	2.87	217
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
1 Private ownership of business should be increased v. 10 Government ownership of business should be increased		0.0138		
2 v. 9		0.0001		
3 v. 8		0.0034		
4 v. 7		0.0223		

Table 22b: Private v. State Ownership of Business and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	765.083	7	109.298	21.201	<0.0001
Within Groups	12,444.674	2,414	5.155		
Total	13,209.757	2,421			

Government Responsibility

Tables 23a and 23b show the results for the government responsibility issue. The rankings show that those who believe that government should take more responsibility (socialist position) are more strongly opposed to tax evasion than those who believe people should take more responsibility (capitalist position). The ANOVA found that the difference between groups was significant at the 1 percent level ($p < 0.0001$).

H23: People are equally averse to tax evasion regardless of attitude toward government vs. individual responsibility.

H23: Rejected.

Table 23a: Ranking by Government Responsibility (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Government Responsibility**	**Mean**	**Std. Dev.**	**n**
1	1 The government should take more responsibility	2.1	2.15	522
1	2	2.1	1.83	337
1	10 People should take more responsibility	2.1	2.38	300
4	3	2.5	1.96	276
4	4	2.5	2.00	205
6	5	2.6	2.12	299
7	8	2.9	2.45	281
7	9	2.9	2.89	190
9	6	3.1	2.46	264
10	7	3.2	2.52	191
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
2 v. 9		0.0001		
3 v. 8		0.0340		
4 v. 7		0.0023		
5 v. 6		0.0098		

Table 23b: Government Responsibility and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	317.747	7	45.392	9.597	<0.0001
Within Groups	11,711.251	2,476	4.730		
Total	12,028.998	2,483			

Confidence in Government

Tables 24a and 24b show the results for the confidence in government question. The results were interesting. Those who have quite a lot of confidence in government and those who do not have confidence in government at all had equal mean scores and both groups were most strongly opposed to tax evasion. The group that was least opposed to tax evasion was the group that had a great deal of confidence in government. The ANOVA found that the difference between groups was significant at the 5 percent level ($p = 0.027$). A t-test comparing the "a great deal" and "quite a lot" groups found the difference in mean score to be significant at the 1 percent level ($p = 0.0038$).

H24: People are equally averse to tax evasion regardless of the degree of confidence in government.
H24: Rejected.

Table 24a: Ranking by Confidence in Government (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Confidence in Government**	**Mean**	**Std. Dev.**	**n**
1	Quite a lot	2.4	2.16	1169
1	Not at all	2.4	2.25	218
3	Not very much	2.5	2.24	616
4	A great deal	2.7	2.48	857
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
A great deal v. Quite a lot		0.0038		

Table 24b: Confidence in Government and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	47.776	3	15.925	3.053	0.027
Within Groups	14,898.550	2,856	5.217		
Total	14,946.326	2,859			

Confidence in the Justice System

Tables 25a and 25b show the results for the category of confidence in the justice system. Those who did not have any confidence in the justice system were most opposed to tax evasion, whereas those who had a great deal of confidence in the justice system were least opposed to tax evasion. The ANOVA found the difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H25: People are equally averse to tax evasion regardless of the degree of confidence in the justice system.
H25: Rejected.

Table 25a: Ranking by Confidence in the Justice System (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Confidence in the Justice System**	**Mean**	**Std. Dev.**	**n**
1	Not at all	2.2	1.86	257
2	Quite a lot	2.4	2.26	1212
2	Not very much	2.4	2.17	700
4	A great deal	2.9	2.51	675
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
A great deal v. Quite a lot		0.0001		
A great deal v. Not very much		0.0001		
A great deal v. Not at all		0.0001		

Table 25b: Confidence in the Justice System and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	150.249	3	50.083	9.736	<0.0001
Within Groups	14,608.750	2,840	5.144		
Total	14,758.999	2,843			

Confidence in the Police

Tables 26a and 26b show the data for the confidence in police question. Those most opposed to tax evasion had quite a lot of confidence in the police, while those least opposed to tax evasion had a great deal of confidence in the police. What this relationship means is unclear. One might conclude a priori that those who have quite a lot or a great deal of confidence in the police would be the two groups that were most strongly opposed to tax evasion, but that was not the case. The ANOVA found the difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H26: People are equally averse to tax evasion regardless of the degree of confidence in the police.
H26: Rejected.

Table 26a: Ranking by Confidence in the Police (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Confidence in the Police**	**Mean**	**Std. Dev.**	**n**
1	Quite a lot	2.3	2.17	1157
2	Not at all	2.4	2.05	372
3	Not very much	2.5	2.34	727
4	A great deal	2.9	2.52	622
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
A great deal v. Quite a lot		0.0001		
A great deal v. Not very much		0.0026		
A great deal v. Not at all		0.0012		

Table 26b: Confidence in the Police and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	149.384	3	49.795	9.591	<0.0001
Within Groups	14,921.500	2,874	5.192		
Total	15,070.884	2,877			

Relative Seriousness of Tax Evasion

The next test was to determine the relative seriousness of tax evasion. The *World Values* surveys collected data on a few other ethical issues. Tables 27a and 27b show the results. Tax evasion was found to be less serious of an offense than accepting a bribe, suicide, or wife beating and more serious than prostitution, claiming government benefits to which you are not entitled, abortion, avoiding a fare on public transport, homosexuality, euthanasia, or divorce. The ANOVA found the difference between groups to be significant at the 1 percent level ($p < 0.0001$).

H27: Tax evasion is equally as serious as other acts.
H27: Rejected.

Table 27a: Ranking by Realtive Seriousness of Tax Evasion (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Seriousness of Tax Evasion**	**Mean**	**Std. Dev.**	**n**
1	Someone accepting a bribe in the course of their duties.	2.4	2.24	2945
1	Suicide	2.4	2.21	2921
1	Wife beating	2.4	2.43	2964
4	Cheating on taxes if you have a chance.	2.5	2.28	2911
4	Prostitution	2.5	2.38	2925
6	Claiming government benefits to which you are not entitled.	2.6	2.51	2931
6	Abortion	2.6	2.53	2910
8	Avoiding a fare on public transport.	2.7	2.41	2932
9	Homosexuality	3.0	2.56	2902
10	Euthanasia	3.2	2.69	2835
11	Divorce	3.9	2.82	2884
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
				p value
Cheating on taxes if you have a chance. v. Avoiding a fare on public transport.				0.0011

Table 27a: Ranking by Realtive Seriousness of Tax Evasion (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)	
Cheating on taxes if you have a chance. v. Homosexuality	0.0001
Cheating on taxes if you have a chance. v. Divorce	0.0001
Cheating on taxes if you have a chance. v. Euthanasia	0.0001
Cheating on taxes if you have a chance. v. Suicide	0.0891
Cheating on taxes if you have a chance. v. Someone accepting a bribe in the course of their duties.	0.0905

Table 27b: Relative Seriousness of Tax Evasion and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	260.464	7	37.209	6.589	<0.0001
Within Groups	132,322.707	23,431	5.647		
Total	132,583.170	23,438			

Trend Analysis

Data on South Africa was gathered in four waves of surveys. A trend analysis was done to determine whether attitude on tax evasion has changed over time. The results in Tables 28a and 28b show that the attitude has changed, but not in a lineal pattern. People were most opposed to tax evasion in the 1996 survey and were least opposed in the 1990 and 2001 surveys. Between 1990 and 2007, opposition increased, then decreased, then increased again. The ANOVA found that the difference between groups was significant at the 1 percent level ($p < 0.0001$). Chart 1 shows the trend.

H28: People are just as averse to tax evasion now as they have been in the past.
H28: Rejected.

Table 28a: Ranking by Trend Analysis (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
Rank	**Trend**	**Mean**	**Std. Dev.**	**n**
1	Wave 3 - 1996	2.2	2.13	2798
2	Wave 5 - 2007	2.5	2.28	2911
3	Wave 2 - 1990	2.7	2.82	2561
3	Wave 4 - 2001	2.7	2.61	2924

Table 28a: Ranking by Trend Analysis (Cheating on taxes is: 1 = never justifiable; 10 = always justifiable)				
SIGNIFICANT DIFFERENCES IN MEAN SCORES				
		p value		
Wave 2 - 1990 v. Wave 3 - 1996		0.0001		
Wave 2 - 1990 v. Wave 5		0.0038		
Wave 3 - 1996 v. Wave 4 - 2001		0.0001		
Wave 3 - 1996 v. Wave 5		0.0001		
Wave 4 - 2001v. Wave 5		0.0018		

Table 28b: Trend Analysis and Attitudes toward Tax Evasion ANOVA Analysis					
	Σ Squares	Df	Mean Squares	Fisher F-value	P value
Between Groups	465.292	3	155.097	25.490	<0.0001
Within Groups	68,086.966	11,190	6.085		
Total	68,552.258	11,193			

CONCLUDING COMMENTS

This study found several interesting relationships between attitude toward tax evasion and more than 20 demographic variables. It is perhaps the most comprehensive demographic

study of South African attitudes toward tax evasion done to date. The methodology used in this study can also serve as a template for studies of other countries and regions. Some of the demographic variables included in this study have not been used in prior studies, which break new ground and may serve as the basis for further research into these variables.

REFERENCES

Adams, C. (1982). *Fight, Flight and Fraud: The Story of Taxation*. Curacao: Euro-Dutch Publishers.

Adams, C. (1993). *For Good or Evil: The Impact of Taxes on the Course of Civilization*. London, New York & Lanham: Madison Books.

Ahmad, M. (1995). *Business Ethics in Islam*. Islamabad, Pakistan: The International Institute of Islamic Thought & The International Institute of Islamic Economics.

Alm, J. & B. Torgler. (2004). Estimating the Determinants of Tax Morale. *Proceedings of the 97th Annual Conference on Taxation of the National Tax Association – Tax Institute of America*, 269-274.

Alm, J., J. Martinez-Vazquez & B. Torgler. (2005). Russian Tax Morale in the 1990s. *Proceedings of the 98th Annual Conference on Taxation of the National Tax Association*, 287-292.

Alm, J. & B. Torgler. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology,* 27, 224-246.

Alm, J., J. Martinez-Vazquez & B. Torgler. (2006). Russian attitudes toward paying taxes – before, during and after the transition. *International Journal of Social Economics,* 33(12), 832-857.

Alm, J., J. Martinez-Vazquez & B. Torgler (2010). Developing Alternative Frameworks for Explaining Tax Compliance. In J. Alm, J. Martinez-Vazquez & B. Torgler (Eds.), *Developing Alternative Frameworks for Explaining Tax Compliance* (pp. 3-12). London & New York: Routledge.

Alm, J. & J. Martinez-Vazquez. (2010). Tax Evasion, the Informal Sector, and Tax Morale in LAC Countries. In J. Alm, J. Martinez-Vazquez & B. Torgler (Eds.), *Developing Alternative Frameworks for Explaining Tax Compliance* (pp. 260-291). London & New York: Routledge.

Armstrong, M.B. & J. Robison. (1998). Ethics in Taxation. *Journal of Accounting, Ethics & Public Policy*, *1*(4), 535-557, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 330-348). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Babakus, E., T. B. Cornwell, V. Mitchell & B. Schlegelmilch (2004). Reactions to Unethical Consumer Behavior across Six Countries. *The Journal of Consumer Marketing*, *21*(4/5), 254-263.

Baird, J.S. (1980). Current Trends in College Cheating. *Psychology in the Schools, 17*(4), 515-522, as cited in B.S. Brown & P. Choong. (2005). An Investigation of Academic Dishonesty among Business Students at Public and Private United States Universities. *International Journal of Management*, *22*(2), 201-214.

Ballas, A. A. & H. Tsoukas. (1998). Consequences of Distrust: The Vicious Circle of Tax Evasion in Greece. *Journal of Accounting, Ethics & Public Policy*, *1*(4), 572-596, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 284-304). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Barnett, J. H. & M. J. Karson. (1987). Personal Values and Business Decisions: An Exploratory Investigation. *Journal of Business Ethics*, 6(5), 371-382.

Barnett, J. H. & M. J. Karson. (1989). Managers, Values, and Executive Decisions: An Exploration of the Role of Gender, Career Stage, Organizational Level, Function, and the Importance of Ethics, Relationships and Results in Managerial Decision-Making. *Journal of Business Ethics*, 8(10), 747-771.

Beito, D.T. (1989). *Taxpayers in Revolt: Tax Resistance during the Great Depression*. Chapel Hill, NC: University of North Carolina Press.

Benk, S., R. W. McGee & A. M. Ross (2009). An Empirical Study of Ethical Opinion in Turkey. *Journal of Accounting, Ethics & Public Policy*, *10*(1), 83-99.

Bird, R. M., J. Martinez-Vazquez & B. Torgler. (2004). Tax Performance in Developing Countries: the Role of Demand Factors. *Proceedings of the 97th Annual Conference on Taxation of the National Tax Association – Tax Institute of America*, 284-289.

Block, W. (1989). The Justification of Taxation in the Public Finance Literature: A Critique. *Journal of Public Finance and Public Choice*, *3,* 141-158.

Block, W. (1993). Public Finance Texts Cannot Justify Government Taxation: A Critique. *Canadian Public Administration/Administration Publique du Canada, 36*(2), 225-262, reprinted in revised form under the title "The Justification for Taxation in the Economics Literature" in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 36-88) The Dumont Institute for Public Policy Research, Dumont, NJ, 1998.

Browning, J. & N. B. Zabriskie. (1983). How Ethical Are Industrial Buyers? *Industrial Marketing Management*, *12*(4), 219-224.

Callan, V.J. (1992). Predicting Ethical Values and Training Needs in Ethics. *Journal of Business Ethics*, *11*(10), 761-769.

Cohn, G. (1998). The Jewish View on Paying Taxes. *Journal of Accounting, Ethics & Public Policy, 1*(2), 109-120, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 180-189). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Cowell, F.A. (1990). *Cheating the Government: The Economics of Evasion*. Cambridge, MA & London: MIT Press.

Crowe, M. T. (1944). The Moral Obligation of Paying Just Taxes, The Catholic University of America Studies in Sacred Theology No. 84.

Cullis, J., P. Jones & A. Lewis (2006). Tax Framing, Instrumentality and Individual Differences: Are There Two Different Cultures? *Journal of Economic Psychology, 27*, 304-320.

Cullis, J., P. Jones & A. Lewis. (2010). Tax Compliance: Social Norms, Culture, and Endogeneity. In J. Alm, J. Martinez-Vazquez & B. Torgler (Eds.), *Developing Alternative Frameworks for Explaining Tax Compliance* (pp. 35-55). London & New York: Routledge.

Cummings, R. G., J. Martinez-Vazquez, M. McKee & B. Torgler. (2004). Effects of Culture on Tax Compliance: A Cross Check of experimental and Survey Evidence. Presented at the Meeting of the Public Choice Society, Baltimore, March 11-14.

Curry, B. (1982). *Principles of Taxation of a Libertarian Society*. Glendale, CA: BC Publishing Company.

DeMoville, W. (1998). The Ethics of Tax Evasion: A Baha'i Perspective. *Journal of Accounting, Ethics & Public Policy*, *1*(3), 356-368, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 230-240). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Gronbacher, G.M.A. 1998. Taxation: Catholic Social Thought and Classical Liberalism. *Journal of Accounting, Ethics & Public Policy, 1*(1), 91-100, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 158-167). Dumont, NJ: The Dumont Institute for Public Policy Research, Dumont, NJ, 1998.

Gupta, R. & R. W. McGee. (2010). A Comparative Study of New Zealanders' Opinion on the Ethics of Tax Evasion: Students v. Accountants. *New Zealand Journal of Taxation Law and Policy, 16*(1), 47-84.

Harris, J. R. (1990). Ethical Values of Individuals at Different Levels in the Organizational Hierarchy of a Single Firm. *Journal of Business Ethics, 9*(9), 741-750.

Izraeli, D. (1988). Ethical Beliefs and Behavior among Managers: A Cross-Cultural Perspective. *Journal of Business Ethics, 7*(4), 263-271.

Jalili, A. R. (2012). The Ethics of Tax Evasion: An Islamic Perspective. In Robert W. McGee (Ed.), *The Ethics of Tax Evasion in Theory and Practice* (forthcoming). New York: Springer.

Johnston, D.C. (2003). *Perfectly Legal: The Covert Campaign to Rig Our Tax System to Benefit the Super Rich – and Cheat Everybody Else*. New York: Penguin.

Johnston, D.C. (2007). *Free Lunch: How the Wealthiest Americans Enrich Themselves at Government Expense (and Stick You with the Bill)*. New York: Penguin.

Kelley, S.W., O.C. Ferrell & S.J. Skinner. (1990). Ethical Behavior Among Marketing Researchers: An Assessment of Selected Demographic Characteristics. *Journal of Business Ethics*, *9*(8), 681-688.

Kidwell, J. M., R. E. Stevens & A. L. Bethke. (1987). Differences in Ethical Perceptions Between Male and Female Managers: Myth or Reality? *Journal of Business Ethics*, *6*(6), 489-493.

Kirchler, E. (2007). *The Economic Psychology of Tax Behaviour*. New York: Cambridge University Press.

Kohut, G. F. & S. E. Corriher. (1994). The Relationship of Age, Gender, Experience and Awareness of Written Ethics Policies to Business Decision Making. *S.A.M. Advanced Management Journal*, *59*(1), 32-39.

Leiker, B. H. (1998). Rousseau and the Legitimacy of Tax Evasion. *Journal of Accounting, Ethics & Public Policy*, *1*(1), 45-57, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 89-101). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Longenecker, J. G., J. A. McKinney & C. W. Moore. (1989). Do Smaller Firms Have Higher Ethics? *Business and Society Review*, *71*, 19-21.

McGee, R. W. (1994). Is Tax Evasion Unethical? *University of Kansas Law Review*, *42*(2), 411-35.

McGee, R. W. (1997). The Ethics of Tax Evasion and Trade Protectionism from an Islamic Perspective, *Commentaries on Law & Public Policy*, *1,* 250-262.

McGee, R.W. (Ed.). (1998a). *The Ethics of Tax Evasion*. Dumont, NJ: The Dumont Institute for Public Policy Research.

McGee, R. W. (1998b). Christian Views on the Ethics of Tax Evasion. *Journal of Accounting, Ethics & Public Policy, 1*(2), 210-225. Reprinted at http://ssrn.com/abstract=461398.

McGee, R. W. (1998c). When Is Tax Evasion Unethical? In Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 5-35). Dumont, NJ: The Dumont Institute.

McGee, R. W. (1998d). The Ethics of Tax Evasion in Islam, in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 214-219). Dumont, NJ: The Dumont Institute.

McGee, R. W. (1998e). The Ethics of Tax Evasion in Islam: A Comment. *Journal of Accounting, Ethics & Public Policy, 1*(2), 162-168, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 214-219). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

McGee, R. W. (1998f). Should Accountants be Punished for Aiding and Abetting Tax Evasion? *Journal of Accounting, Ethics & Public Policy, 1*(1), 16-44.

McGee, R. W. (1998g). Jewish Views on the Ethics of Tax Evasion, *Journal of Accounting, Ethics & Public Policy, 1*(3), 323-336.

McGee, R. W. (1999a). Is It Unethical to Evade Taxes in an Evil or Corrupt State? A Look at Jewish, Christian, Muslim, Mormon and Baha'i Perspectives. *Journal of Accounting, Ethics & Public Policy*, *2*(1), 149-181. Reprinted at http://ssrn.com/abstract=251469.

McGee, R. W. (1999b). An Ethical Look at Paying Your "Fair Share" of Taxes. *Journal of Accounting, Ethics & Public Policy*, *2*(2), 318-328. Reprinted at http://ssrn.com/abstract=242549.

McGee, R. W. (1999c). Is It Unethical to Evade the Estate Tax? *Journal of Accounting, Ethics & Public Policy*, *2*(2), 266-285. Reprinted at http://ssrn.com/abstract=242536.

McGee, R. W. (1999d). Tariffs as a Form of Taxation: Is Evasion Unethical? *Journal of Accounting, Ethics & Public Policy*, *2*(2), 376-385. Reprinted at http://ssrn.com/abstract=242565.

McGee, R. W. (1999e). Why People Evade Taxes in Armenia: A Look at an Ethical Issue Based on a Summary of Interviews. *Journal of Accounting, Ethics & Public Policy*, *2*(2), 408-416. Reprinted at http://ssrn.com/abstract=242568.

McGee, R. W. (2004). *The Philosophy of Taxation and Public Finance*. Norwell, MA and Dordrecht: Kluwer Academic Publishers.

McGee, R. W. (2006a). Three Views on the Ethics of Tax Evasion. *Journal of Business Ethics*, *67*(1), 15-35.

McGee, R. W. (2006b). A Survey of International Business Academics on the Ethics of Tax Evasion. *Journal of Accounting, Ethics & Public Policy*, *6*(3), 301-352.

McGee, R. W. (2006c). The Ethics of Tax Evasion: A Survey of Romanian Business Students and Faculty. *The ICFAI Journal of Public Finance*, *4*(2), 38-68. Reprinted in Robert W. McGee & Galina G. Preobragenskaya, *Accounting and Financial System Reform in Eastern Europe and Asia* (pp. 299-334). New York: Springer, 2006.

McGee, R. W. & A. Bernal (2006). The Ethics of Tax Evasion: A Survey of Business Students in Poland. In Mina Baliamoune-Lutz, Alojzy Z. Nowak & Jeff Steagall (Eds.), *Global Economy -- How It Works* (pp. 155-174). Warsaw: University of Warsaw & Jacksonville: University of North Florida. Reprinted at http://ssrn.com/abstract=875434.

McGee, R. W. & T. Maranjyan (2006). Tax Evasion in Armenia: An Empirical Study. Working Paper No. 06/10, Armenian International Policy Research Group, Washington, DC.

McGee, R. W., I. Nickerson & W. Fees (2006). German and American Opinion on the Ethics of Tax Evasion. *Proceedings of the Academy of Legal, Ethical and Regulatory Issues* (Reno), *10*(2), 31-34.

McGee, R. W. & Z. Guo (2007). A Survey of Law, Business and Philosophy Students in China on the Ethics of Tax Evasion. *Society and Business Review*, *2*(3), 299-315.

McGee, R. W., C. Noronha & M. Tyler (2007). The Ethics of Tax Evasion: a Survey of Macau Opinion. *Euro Asia Journal of Management*, 17(2), 123-150. Reprinted in Robert W. McGee (Ed.), *Readings in Accounting Ethics* (pp. 283-313), Hyderabad, India: ICFAI University Press, 2009.

McGee, R. W. & S. López Paláu (2007). The Ethics of Tax Evasion: Two Empirical Studies of Puerto Rican Opinion. *Journal of Applied Business and Economics, 7*(3), 27-47. Reprinted in Robert W. McGee (Ed.), *Readings in Accounting Ethics* (pp. 314-342). Hyderabad, India: ICFAI University Press, 2009.

McGee, R. W. (2008). Opinions on Tax Evasion in Thailand, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 609-620). New York: Springer.

McGee, R. W. (2008a). A Survey of Vietnamese Opinion on the Ethics of Tax Evasion, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 663-674). New York: Springer.

McGee, R. W., J. Alver & L. Alver (2008). The Ethics of Tax Evasion: A Survey of Estonian Opinion, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 461-480). New York: Springer.

McGee, R. W. & Y. An (2008). A Survey of Chinese Business and Economics Students on the Ethics of Tax Evasion, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 409-421). New York: Springer.

McGee, R. & Y.Butt (2008). An Empirical Study of Tax Evasion Ethics in Hong Kong. Proceedings of the International Academy of Business and Public Administration Disciplines , Dallas, April 24-27, 72-83.

McGee, R W. & G. M. Cohn (2008). Jewish Perspectives on the Ethics of Tax Evasion. *Journal of Legal, Ethical and Regulatory Issues*, *11*(2), 1-32.

McGee, R. W., S.S.M. Ho & A.Y.S. Li (2008). A Comparative Study on Perceived Ethics of Tax Evasion: Hong Kong vs. the United States. *Journal of Business Ethics*, *77*(2), 147-158.

McGee, R. W. & C. Lingle (2008). The Ethics of Tax Evasion: A Survey of Guatemalan Opinion, in R.W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 481-495). New York: Springer.

McGee, R. W. & S. López Paláu (2008). Tax Evasion and Ethics: A Comparative Study of the USA and Four Latin American Countries, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 185-224). New York: Springer.

McGee, R. W. & B. M'Zali (2008). Attitudes toward Tax Evasion in Mali, in Robert W. McGee, editor, *Taxation and Public Finance in Transition and Developing Economies* (pp. 511-517). New York: Springer.

McGee, R. W. & C. Noronha (2008). The Ethics of Tax Evasion: A Comparative Study of Guangzhou (Southern China) and Macau Opinions. *Euro Asia Journal of Management*, *18*(2), 133-152.

McGee, R. & G.Preobragenskaya (2008). A Study of Tax Evasion Ethics in Kazakhstan, in R. W. McGee, editor, *Taxation and Public Finance in Transition and Developing Economies* (pp. 497-510). New York: Springer.

McGee, R. W. & M. J. Rossi (2008). A Survey of Argentina on the Ethics of Tax Evasion, in R.W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 239-261). New York: Springer.

McGee, R. W. & R. Tusan (2008). The Ethics of Tax Evasion: A Survey of Slovak Opinion, in R.W. McGee, editor, *Taxation and Public Finance in Transition and Developing Economies* (pp. 575-601). New York: Springer.

McGee, R. W. (2009). Views toward Tax Evasion: A Comparative Study of Moldova and Romania. *ICFAI University Journal of Public Finance*, *7*(3&4), 7-24.

McGee, R. W. & S. N. Vittadini Andres (2009). The Ethics of Tax Evasion: Case Studies of Taiwan, in Robert W. McGee, *Readings in Business Ethics* (pp. 200-228). Hyderabad, India: ICFAI University Press. An abbreviated version was published in Marjorie G. Adams and Abbass Alkhafaji (Ed.), *Business Research Yearbook: Global Business Perspectives*, Volume XIV, No. 1 (pp. 34-39). Beltsville, MD: International Graphics: Beltsville, MD, 2007.

McGee, R. W., M. Basic & M. Tyler (2009). The Ethics of Tax Evasion: A Survey of Bosnian Opinion, *Journal of Balkan and Near Eastern Studies*, *11*(2), 197-207.

McGee, R. W., S. Benk, A. M. Ross & H. Kılıçaslan (2009). An Empirical Study of Ethical Opinion in Germany. *Journal of Accounting, Ethics & Public Policy*, *10*(2), 243-259.

McGee, R. W. & S. Bose (2009). The Ethics of Tax Evasion: A Survey of Australian Opinion, in Robert W. McGee, *Readings in Business Ethics* (pp. 143-166). Hyderabad, India: ICFAI University Press.

McGee, R. W. & B. M'Zali (2009). The Ethics of Tax Evasion: An Empirical Study of French EMBA Students, in Robert W. McGee, *Readings in Business Ethics* (pp. 185-199). Hyderabad, India: ICFAI University Press. An abbreviated version was published in Marjorie G. Adams and Abbass Alkhafaji (Eds.), *Business Research Yearbook: Global Business Perspectives*, Volume XIV, No. 1 (pp. 27-33). Beltsville, MD: International Graphics, 2007).

McGee, R. W., I. Nickerson & W. Fees (2009). When Is Tax Evasion Ethically Justifiable? A Survey of German Opinion, in Robert W. McGee (Ed.), *Readings in Accounting Ethics* (pp. 365-389). Hyderabad, India: ICFAI University Press.

McGee, R. W., S. López Paláu & G. A. Yepes Lopez (2009). The Ethics of Tax Evasion: An Empirical Study of Colombian Opinion, in Robert W. McGee, *Readings in Business Ethics* (pp. 167-184). Hyderabad, India: ICFAI University Press.

McGee, R. & S.Smith (2009). Ethics and Tax Evasion: A Comparative Study of Utah and Florida Opinion, in R.W. McGee (Ed.), *Readings in Accounting Ethics* (pp. 343-364). Hyderabad, India: ICFAI University Press.

McGee, R.W. & G.A. Goldman (2010). Ethics and Tax Evasion: A Survey of South African Opinion. *Proceedings of the Third Annual University of Johannesburg Faculty of Management Conference*, May 12-14.

McGee, R. W. & S. Benk (2011). *The Ethics of Tax Evasion: A Study of Turkish Opinion. Journal of Balkan and Near Eastern Studies*, forthcoming.

McGee, R. W., S. Benk, H. Yıldırım & M. Kayıkçı. (2011). The Ethics of Tax Evasion: A Study of Turkish Tax Practitioner Opinion, *European Journal of Social Sciences*, 18(3), 468-480.

McGee, R. W. (Ed.) (2012). *The Ethics of Tax Evasion in Theory and Practice*. New York: Springer.

Morales, A. (1998). Income Tax Compliance and Alternative Views of Ethics and Human Nature. *Journal of Accounting, Ethics & Public Policy*, 1(3), 380-399, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 242-258). The Dumont Institute for Public Policy Research: Dumont, NJ, 1998.

Murtuza, A. & S.M. Ghazanfar (1998). Taxation as a Form of Worship: Exploring the Nature of Zakat. *Journal of Accounting, Ethics & Public Policy*, *1*(2), 134-161, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 190-212). The Dumont Institute for Public Policy Research: Dumont, NJ, 1998.

Nasadyuk, I. & R. W. McGee (2008). The Ethics of Tax Evasion: An Empirical Study of Business and Economics Student Opinion in Ukraine, in Robert W. McGee (Ed.), *Taxation and Public Finance in Transition and Developing Economies* (pp. 639-661). New York: Springer.

Nozick, R. (1974). *Anarchy, State & Utopia.* New York: Basic Books.

Oliva, R. R. (1998). The Schism between Tax Practitioners' Ethical and Legal Obligations: Recommendations for the Fusion of Law and Ethics. *Journal of Accounting, Ethics & Public Policy*, *1*(4), 603-628, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 350-371). Dumont, NJ: The Dumont Institute for Public Policy Research (1998).

Pennock, R. T. (1998). Death and Taxes: On the Justice of Conscientious War Tax Resistance. *Journal of Accounting, Ethics & Public Policy*, 1(1), 58-76, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 124-142). Dumont, NJ: The Dumont Institute for Public Policy Research (1998).

Preobragenskaya, G. G. & R. W. McGee (2004). Taxation and Public Finance in a Transition Economy: A Case Study of Russia. In Carolyn Gardner, Jerry Biberman & Abbass Alkhafaji (Eds.), *Business Research Yearbook*: *Global Business Perspectives* Volume XI (pp. 254-258), Saline, MI: McNaughton & Gunn, Inc. (2004). A longer version, which was presented at the Sixteenth Annual Conference of the International Academy of Business Disciplines in San Antonio, March 25-28, 2004, is available at http://ssrn.com/abstract=480862.

Ruegger, D. & E. W. King (1992). A Study of the Effect of Age and Gender upon Student Business Ethics. *Journal of Business Ethics*, *11*(3), 179-186.

Sabrin, M. (1995). *Tax Free 2000: The Rebirth of American Liberty*. Lafayette, LA: Prescott Press.

Schansberg, D. Eric (1998). The Ethics of Tax Evasion within Biblical Christianity: Are There Limits to "Rendering Unto Caesar"? *Journal of Accounting, Ethics & Public Policy*, *1*(1), 77-90, reprinted in Robert W. McGee (Ed.), The Ethics of Tax Evasion (pp. 144-157). The Dumont Institute for Public Policy Research,

Serwinek, P. J. (1992). Demographic & Related Differences in Ethical Views Among Small Businesses. *Journal of Business Ethics*, 11(7), 555-566.

Sims, R. R., H. K. Cheng & H. Teegen. (1996). Toward a Profile of Student Software Piraters. *Journal of Business Ethics*, 15(8), 839-849.

Smatrakalev, G. (1998). Walking on the Edge: Bulgaria and the Transition to a Market Economy. In R.W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 316-329) Dumont, NJ: The Dumont Institute for Public Policy Research.

Smith, S. R. & K. C. Kimball (1998). Tax Evasion and Ethics: A Perspective from Members of The Church of Jesus Christ of Latter-Day Saints. *Journal of Accounting, Ethics & Public Policy*, *1*(3), 337-348, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 220-229). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Tamari, M. (1998). Ethical Issues in Tax Evasion: A Jewish Perspective. *Journal of Accounting, Ethics & Public Policy, 1*(2), 121-132, reprinted in Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 168-178). Dumont, NJ: The Dumont Institute for Public Policy Research, 1998.

Thorndike, J.& D.Ventry, Jr. (2002). *The Ongoing Debate on Tax Justice*. Washington, DC: The Urban Institute Press.

Torgler, B. (2003a). Tax Morale: Theory and Empirical Analysis of Tax Compliance. Dissertation der Universität Basel zur Erlangung der Würde eines Doktors der Staatswissenschaften.

Torgler, B. (2003b). Tax Morale in Transition Countries. *Post-Communist Economies, 15*(3), 357-381.

Torgler, B. (2003c). Tax Morale and Tax Compliance: A Cross Culture Comparison. *Proceedings of the 96th Annual Conference on Taxation of the National Tax Association-Tax Institute of America*, 63-74.

Torgler, B. (2007). *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*. Cheltenham, UK & Northampton, MA, USA: Edward Elgar.

Vaguine, V. V. (1998). The "Shadow Economy" and Tax Evasion in Russia. In Robert W. McGee (Ed.), *The Ethics of Tax Evasion* (pp. 306-314) Dumont, NJ: The Dumont Institute for Public Policy Research.

Webber, C. & A. Wildavsky. (1986). *A History of Taxation and Expenditure in the Western World*. New York: Simon & Schuster.

World Values Survey Data. www.wvsevsdb.com/

Yusuf, S.M. (1971). *Economic Justice in Islam*. Lahore: Sh. Muhammad Ashraf.

ECONOMETRIC STUDY OF TIME USE AND SCORES IN ONLINE MBA STATISTICS CLASS: A GENDER ANALYSIS

Kishor K. Guru-Gharana, Texas A & M University-Commerce
Jennifer Flanagan, Texas A & M University-Commerce

ABSTRACT

Improved Student learning is the ultimate goal of educators and is generally measured in terms of scores earned in the course. Students themselves must also dedicate adequate hours to the course. The present study provides evidence that a student's final grade is closely linked to the hours spent in the course, especially with regard to online statistic courses. This study uses actual recorded online time use of students instead of self-reported surveys used in most studies in the relevant literature. Moreover, the models use actual scores instead of the letter grades which hide a lot of information by converting the ratio scale variable to discrete ordinal variable. As a result, this study could use Constant elasticity and Decreasing Elasticity Mixed Dummy Multiple Regression models assuming that online time use is an objectively measurable good indicator of overall effort by students in online classes. The evidences suggest that there is a significant reward for additional effort, especially at the lower levels of times use and scores. The Constant Elasticity model predicts a 4.3% improvement in existing score for additional 10% increase in online time use for male students. For female students the improvement is expected to be only about 2.5% in existing score. The gender difference is highly significant statistically in the Constant Elasticity model. The decreasing Elasticity model is not only theoretically more appealing but also most successful in explaining variations in the scores, although the gender difference gets dampened and loses some of its statistical significance in this model. According to this model, a 10% increase in online time use for male students with minimal online time use (about 9.2 hours over the semester), is expected to improve the existing score by 3.8% of existing score. For a similar female student the predicted improvement is 3.1% of existing score. As the level of time use increases to the mean level (76.4 hours over the semester), the elasticity for male students drops to 0.05 indicating that a 10% increase in time use would be expected to improve existing score only by 0.5%. The gender difference at higher levels of time use becomes very small. The results of this study are particularly significant for students with low online time use. Instructors should encourage such students to significantly increase their effort as it promises much larger reward at the lower end of time use. Although few students can and have achieved high scores despite their low online time use, it is clear from the data that very low online time use is a good predictor of low scores with few exceptions. This research can be extended by including other objectively measurable attributes and also covering other subjects.

INTRODUCTION

Online MBA is relatively a recent but exponentially expanding phenomenon. The factors contributing to students' success in this online paradigm is topic of recurrent interest for Higher Education Administrators, Academicians, Communities and Students. There have been several studies in the relevant literature which attempt to identify and quantify the relationships of various factors with students' academic performance/achievement. But there is a dearth of econometric study of the impact of the single most important key factor, namely, students' effort measured by online time use, on students' performance with a gender perspective, especially for graduate students. In the literature pertaining to student participation and effort, most studies have concentrated on the simple measure of attendance. Many studies have found that class attendance positively affects performance across various subjects. For example, Devadoss and Foltz (1996) for Agricultural Economics; Schmidt (1983), Park and Kerr (1990), Romer (1993), Durden and Ellis (1995), Ellis and Durden (1998) and Cohn and Johnson (2006) for Economics; Chan et al. (1997) and Johnson et al. (2002) for Finance; Gunn (1993) and Launius (1997) for Psychology; Day (1994) for Sociology; Gump (2005) for General Education; Ledman and Kamuche (2002) for Statistics; Rodgers (2001 and 2002) for Business; Biology Gatherer and Manning (1998) for Biology; and Nyamapfene (2010) for Electronic Engineering. In contrast some studies, such as Buckles and McMahon (1971) and Douglas and Sulock (1994) found no significant contribution of class attendance on performance.

These studies provide evidence of importance of attendance for students in conventional face-to-face courses. However, the question for online and distance courses, where there is often no requirement to attend at a particular place or at a particular time is different. Since class attendance is not relevant some other measure has to be used. In online courses almost all of the interaction is through online tools such as Virtual Office Forum, Students' Forum, Discussion Board, emails to the Instructor and to peers, downloading materials posted by the Instructor, Class live sessions, etc. Participation can be defined by activities such as logging in to symmetric class sessions, participation in discussion boards and forums, online interaction with peers and Instructor, and downloading and reading class materials. There have been relatively few studies on online courses with respect to students' efforts and their performance.

This is what we propose to do in this study. The present study differs from the prevalent literature in its empirical strategy and information content of the data. We use Multiple Regression with mixed Dummy models for Gender with a data consisting of the actual scores instead of data classified into letter grades. The actual score is a ratio scale quantitative variable while letter grade is an ordinal scale discrete variable. The information content is significantly different. For example, a score of 800 is treated equally as a score of 899 (letter grade B). Obviously, the measurement of marginal impact will not only be dramatically blurred but also distorted in a very nonlinear and asymmetric way when scores are converted into letter grades. For example, a change of score from 800 to 895 will not be noticed while a much smaller change

from 895 to 900 will be recorded as change of letter grade from "B" to "A". Consequently, observations with scores marginally below some letter grade will have relatively much larger influence on positive effects; those with scores marginally above some letter grade will have much larger influence on negative effects; and those with scores in the middle will have much weaker influence either way. Thus, all observations will not be treated equally as assumed by most statistical models used in this case. Clearly, the information extracted afterwards will have severe distortions whatsoever elegant (and high sounding) statistical model may a researcher apply. Even a subsequent conversion of letter grades to a quantitative variable like GPA does not bring back the information already lost and repair the asymmetric problem mentioned above. Moreover, the set of econometric tools which can be applied to ordinal dependent variable instead of quantitative also becomes quite restricted with several limitations. Ordinary Multiple Regressions become inappropriate and models such as Multinomial Logit have to be resorted to.

Therefore, the present study makes an important contribution by using quantitative measurement of students' performance, using Multiple Regressions and Mixed Dummy (Slope and Intercept) for Gender. It is often useful to employ a methodology different from those in prior studies to provide a fresh perspective and confirmation or contradiction of previous findings.

Another important deviation of the present study from the prevalent literature is in the collection of data. Many studies rely on the data collected from students' surveys where the students do self-reporting. Such data clearly involve large measurement errors whose size and nature are not easy to estimate. This approach clearly lacks desired accuracy and completeness of the collected data. Stinebrickner and Stinebrickner (2004) emphasize that the reporting error from retrospective survey questions is likely to be substantial. They discuss estimators that might be appropriate when reporting errors are common yet highlight the limitations of the results obtained from the analysis of such data samples.

Our data consist of actually recorded time use and scores in the e-college system used by online courses of Texas A & M University-Commerce, a medium size public University with AACSB accredited online MBA. Our sample includes 308 students who completed graduate level Statistics course in the MBA program from Fall of 2009 to Fall of 2011. The e-college system keeps a detailed record of individual student activity with a precise measure of the time (in minutes) each student spends on each activity of a course. The rest of the paper is organized as follows. Section 2 contains a brief review of the relevant literature. Section 3 contains a description of our data sample and econometric methodology. Section 4 presents the empirical results and section 5 concludes.

REVIEW OF SELECTED LITERATURE

There have been several studies on class attendance and students' performance in traditional face-to-face classes. Schmidt (1983) reported that time spent attending lectures in a

macroeconomic principles course contributed positively to performance. Park and Kerr (1990) found that attendance was a determinant of student performance in a money and banking course, although it was not as important as a student's GPA and percentile rank on a college entrance exam. Romer (1993) found that attendance did contribute significantly to the academic performance of students in a large intermediate macroeconomics course. The author admits the major problem with the study because attendance was not taken every day, thus involving a statistical measurement error. Devadoss and Foltz (1995) found, for a sample of students enrolled in agricultural economics and agribusiness courses, that the more classes attended, the better the students' grades. Durden and Ellis (1995) found that student absences had a significant, negative effect on student performance in the principles of economics course. They find a nonlinear relationship inferring that a few absences do not impact grades, but more than four were found to negatively impact grades.

Ellis, et al. (1998) use data collected by surveying students at **Appalachian State University** at the end of the semester in several sections of the principles of economics course (both micro and macro). A questionnaire was administered over five semesters: Spring and Fall 1993, Spring and Fall 1994, and Spring 1995. The data on absences were estimated number of classes missed during the semester as reported by the students themselves. The observations on student grades were simply the percentage of possible course points earned by the student for the semester. This study treats student performance as a dichotomous variable considering the student as either having done well or having done poorly in the course. The logit analysis employed in this study showed that the probability of a student earning a grade of A or B in Principles of Economics declines as the number of missed classes increases, and the probability of a student earning a D or F increases as classes missed increases. Other factors that positively affect the chances of earning a good grade are the student's GPA, taking calculus and SAT scores. Other negative influences include membership in a fraternity or sorority and the number of credit hours carried during the semester. This study also finds that while females are just as likely as males to do well in principles, they are more likely than males to do poorly for virtually all levels of class attendance, other things being held constant.

Burrus et al. (2001) find that hours of study and student perception concerning the usefulness of homework assignments in preparing for exams increases a student's performance on homework assignments. Ninety-eight students in Principles of Macroeconomics, a prerequisite for all business majors, are surveyed about their perception of homework effectiveness during the Spring and Fall semesters of 1999.2 Students provide categorical information on their GPA's (*GPA)*, hours spent studying course material (*HSD)*, the perceived usefulness (*USE*) of the homework in exam-preparation and the time (*TME*) given to complete the assignments.3 Students complete the surveys during the final class meeting. Marburger (2001) uses detailed information on 60 students enrolled in a section of microeconomics principles over a single semester to investigate the impact of attendance on particular days on exam grades. In his study, lecture material is matched with respective multiple choices questions

to determine if a student is more likely to miss a question covered on the day of an absence. In his study, lecture material is matched with respective multiple choices questions to determine if a student is more likely to miss a question covered on the day of an absence. Ledman and Kamuche (2002) using correlation analysis and tests of hypothesis show that student test performance is better when class attendance is better and that students with better attendance demonstrate more knowledge of the course material. Johnson et al. (2002) study the relation between performance and effort by students in an introductory financial management course. Instead of relying on self-reported data, this study used objectively measured data on effort by the number of attempts made and the amount of time spent by students on repeatable computerized quizzes. The authors find that effort positively influences student performance and encourage educators to motivate students to exert effort in their education. Stinebrickner and Stinebrickner (2004) study the relationship between educational outcomes and students' study time and effort using unique new data from the Berea Panel Study.

Brookshire and Palocsay (2005) analyze the performance of undergraduate students in management science courses and report that overall academic achievement as measured by students' GPA has a significantly higher impact on achievement than students' mathematical skills as measured by math SAT scores. The study by Lin and Chen (2006) considers the effect of cumulative class attendance while estimating the relationship between class attendance and students' exam performance, using an individual-level data. They find that, cumulative attendance produces a positive and significant impact on students' exam performance. Attending lectures corresponds to a 4% improvement in exam performance, and the marginal impact of cumulative attendance on exam performance is also close to 4%. However, the impact of attendance on exam performance is reduced about 0.4% after controlling for the cumulative attendance effect. Cohn and Johnson (2006) use a sample of 347 students, enrolled in principles of economics classes during the period 1997-2001 to examine the relation between class attendance and student performance on examinations. Marburger (2006) investigates the impact of enforcing an attendance policy on absenteeism and student performance and concludes that an enforced mandatory attendance policy significantly reduces absenteeism and improves exam performance. Stinebrickner and Stinebrickner (2008) examine the causal effect of studying on grade performance using an Instrumental Variable estimator using longitudinal data and suggest that human capital accumulation is far from predetermined at the time of college entrance.

More recently, Crede et al. (2010) find class attendance as a better predictor of college grades than any other known predictor of college grades—including SAT scores, HSGPA, studying skills, and the amount of time spent studying. They conclude that the relationship is so strong as to suggest that dramatic improvements in average grades (and failure rates) could be achieved by efforts to increase class attendance rates among college students. Nyamapfene (2010) studies the impact of class attendance on academic performance in a second year Electronics Engineering course module with online notes and no mandatory class attendance policy. The study shows that class attendance is highly correlated to academic performance,

despite the availability of online class notes. In addition, there is significant correlation between class attendance and non-class contact with the lecturer and between student performance in the first year of university study and current academic performance and class attendance. However, there is no correlation between pre-university academic performance and current class attendance and academic performance. The study finds no gender bias in either class attendance or academic performance. Lastly, the study finds that a student's choice of degree program has no impact on class attendance and academic performance in this particular course module.

The literature on the relation between students' time spent and performance is relatively new and smaller, and most studies again relay on self reported surveys. Nonis et al. (2005) analyze survey data containing demographic, behavioral, and personality variables of 228 undergraduate students attending a medium size AACSB accredited public university. Using a hierarchical regression model they find that self-reported time per credit hour spent on academic activities outside of class explains a significant portion of the variation in the semester grade point average (GPA) for senior students, but has no impact on the cumulative GPA. George et al.(2008) use a sample of 231 students attending a private liberal arts university in central Alberta, Canada, who completed a 5-day time diary and a 71-item questionnaire assessing the influence of personal, cognitive, and attitudinal factors on success. The authors find that the greatest predictors of GPA were time-management skills, intelligence, time spent studying, computer ownership, less time spent in passive leisure, and a healthy diet.

Brint et al. (2010) find that there is a surprisingly modest relationship between UC GPA and reported hours studying reflecting differences in academic requirements and perhaps grading practices across disciplines as well as differences in individual effort required to obtain a given level of performance. On the other hand, they find that high school GPA is a good predictor of time spent studying. Students with stronger high school GPAs studied more at UC than those who had lower high school GPA. Brint and Cantwell (2010) use survey of about 6000 responses to the 2006 University of California Undergraduate Experience Survey (UCCUES). Controlling for students' socio-demographic backgrounds, previous academic achievements, and social psychological stressors, they find that study time is strongly connected to both academic conscientiousness and higher grade point averages. Uses of time that connect students to campus life showed relatively weak and inconsistent effects. This study suggests stronger focus on academic study time as the central key to positive academic outcomes, and a renewed focus on off-campus work as a major obstacle to positive academic outcomes.

In contrast to the large and growing literature related to face-to-face education there have been relatively few studies on the relationship between students' participation and performance in distance learning classes, especially at the graduate level. Cheung and Kan (2002) reported on students enrolled in a business course at the Open University of Hong Kong; they found a relationship between tutorial attendance and performance in a hybrid course. Riffell and Sibley (2005) evaluated the effectiveness of the online portion of a hybrid course in an introductory environmental biology course for non-science majors and found no relationship between lecture

attendance and post-test scores. Picciano (2002) studied a totally asynchronous online graduate education administration course, dividing the students into three groups by level of participation. He found no difference in exam performance, but the high participation group (measured by substantial discussion board posts) scored significantly better on the written assignments.

Douglas and Alemanne (2007) present data comparing measures of student effort with student success on an online course. The course which is part of an online Masters program in Library and Information Science ran in the spring semester of 2007. Data was collected from 30 of the 32 students on the course. Participation was measured by counting discussion posts, class utterances, email contacts and course web site clicks. The authors conclude that class participation, no matter how crudely measured, is an important factor in academic success. However this study suffers from the limitation of the small sample size and the accuracy of the measurement of participation. For example, the students could be clicking just to improve their click count and overall grade which was partially based on such participation. Damianov et al. (2009) examine the determinants of academic achievement in online business courses. As a measure of effort, this study uses the total amount of time each student spent in the course. This study estimates a multinomial logistic model to examine the odds of attaining one grade versus another depending on time spent online, GPA, and some demographic characteristics of students. This study finds that extra effort can help a student move from letter grades F, D and C to grade B, but is less helpful for the move from B to A. For the latter improvement, a high GPA matters the most.

DATA AND METHODOLOGY

The present study is based on the recorded online time use and score of 308 MBA students in online graduate Statistics classes taught by the same Instructor at Texas A & M University- Commerce, a midsize public Institution with AACSB accredited online MBA.

This course is a core course with application across other courses and general observation suggests that achievement in this course is highly correlated with overall achievement in the program. The e-college system keeps a detailed record of individual student activity with a precise measure of the time (in minutes) each student spends on each activity of a course. The observations were classified by gender based on names and utmost care was taken for accuracy. After gender classification the names were removed to make the data completely without any identifier. The variables used in this study are Online Time Use in minutes denoted by “Tm”, actual scores out of the total 1000 denoted by “Sc”, and Gender denoted by “Ge”. The value assigned is 0 for Male and 1 for female. Thus Gender is a Dummy variable with Male as the base case. There were 166 male and 142 female students in the sample of 308. A basic assumption underlying this study is that online time use is a good objectively measurable indicator of overall effort by students in online classes. There may be some exceptions, but these two seem to be highly correlated in general.

We use three types of Mixed Dummy Multiple regression models:

Type1: Linear Multiple Regression

$\widehat{Sc} = b_0 + b_1 Tm + b_2 Ge*Tm + b_3 Ge$ (1)

Type2: Constant Elasticity Double Log Multiple Regression

$\widehat{\log(Sc)} = b_0 + b_1 \log(Tm) + b_2 Ge*\log(Tm) + b_3 Ge$ (2)

Type3: Decreasing Elasticity Linear-Log Multiple Regression

$\widehat{Sc} = b_0 + b_1 \log(Tm) + b_2 Ge*\log(Tm) + b_3 Ge$ (3)

The discussion about the properties of the different functional forms and the interpretations of marginal effects and elasticity can be found in any standard Econometrics book, such as Asteriou and Hall (2007, pages 161-65). The slope coefficient b_1 of the first model provides the estimated marginal impact of one unit change in time use (in minutes) on score for male students (out of 1000), while $b_1 + b_2$ provides the marginal impact of one unit change in time use on score for female students. The constant term b_0 is the estimated intercept for male students. In other words, it is the expected score (out of 1000) when online time use is zero (or nearly zero for practical purposes) for male students. On the other hand, $b_0 + b_3$ is the estimated intercept corresponding to female students.

The second model uses score and time in their (natural) log forms. This model measures elasticity (unit free) instead of marginal impact. Here b_1 measures the elasticity of scores with respect to time use for male students, that is, the percentage change is score as a result of percentage change in time use. Similarly, $b_1 + b_2$ measures the same for female students. The anti-log of the constant term is simply the scale factor. Since the scatter plot of score and time use showed somewhat nonlinear relation we tried the log-linear model along with linear model. We compare various aspects of these two models including overall explanatory power, significance of coefficients and various selection criteria. We also report several econometric tests.

Theoretically, the more appealing model is the one with variable elasticity. Here the elasticity is equal to b_1/Sc and the marginal effect is b_1/Tm. The elasticity and marginal effect decline as the levels of score and time use increase. This makes sense if we expect it to be more challenging to improve the score with effort only as the level of score increases. There is a general perception that it is harder to jump from B to A than from C to B on the basis of effort only, as also confirmed by the Multinomial Logit model findings of Damianov et al. (2009). Thus, we are also able to show that you don't need to resort to Multinomial Logit model, as perhaps claimed by Damianov et al. (2009), for estimating such asymmetric effects of efforts on achievement.

One can also argue that the variation in elasticity could be just the reverse, considering the fact that it is generally harder to earn the first million (or billion) than to earn the next million (or billion): a self reinforcing process as the student does better and better with more and more efforts. To check this, a fourth model with log of Score but Time use in original form was tried, but the results were relatively poor. All estimations and tests were done using 7th edition of EVIEWS.

EMPIRICAL RESULTS

Descriptive Statistics

The distributions of scores are displayed in figures 1, 2 and 3 below. The overall mean score is 850.3, male mean score is 845.2 and female mean score is 856.3. Thus, the average of female student score is about 1.1 percentage point higher than male student. The male students have relatively much higher variation with a standard deviation of about 120/1000 compared to only about 85/1000 for female students. However, a test of difference between the mean scores in Table 1 indicates that the difference is statistically quite insignificant.

Figure 1: Distribution of Scores (Pooled Data)

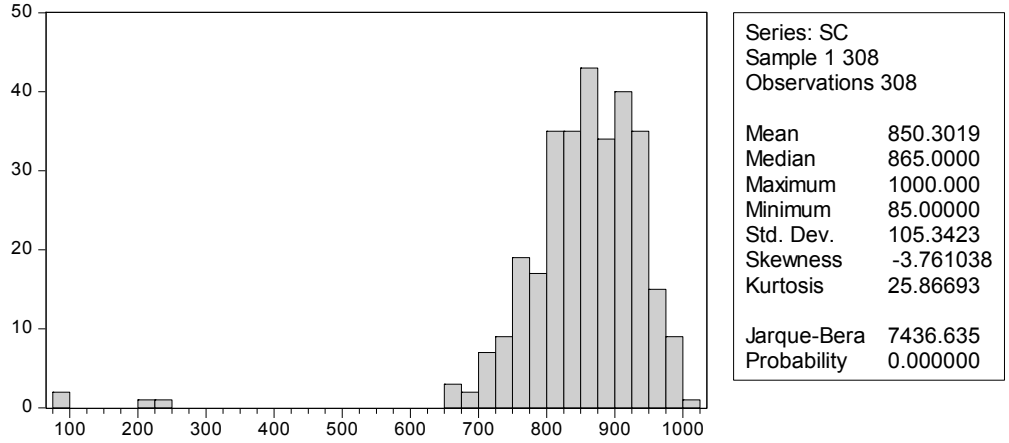


Figure 2: Distribution of Scores (Male)

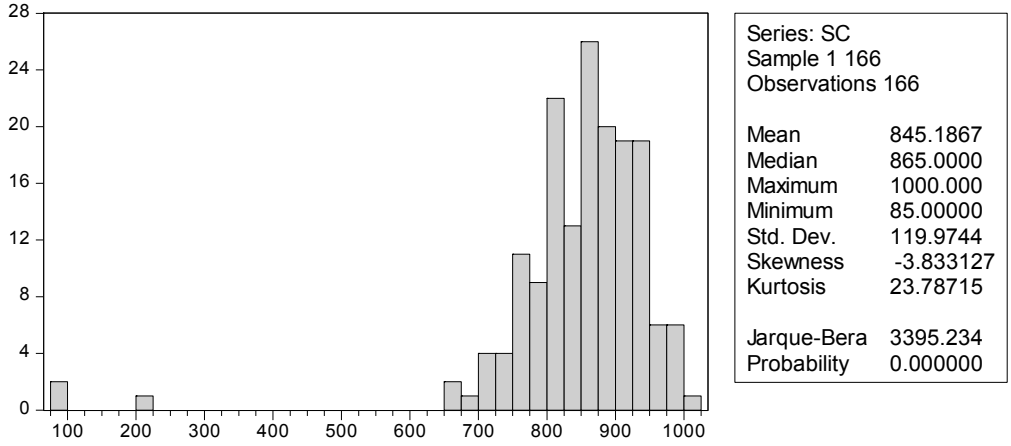


Figure 3: Distribution of Scores (Female)

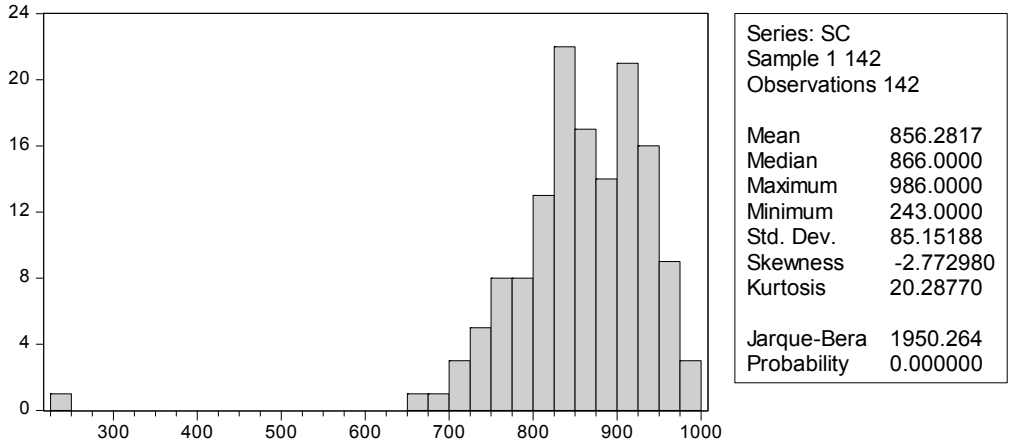


Table 1: Test of Difference between Mean Scores			
Hypothesis Test: Independent Groups (t-test, unequal variance)			
	maleSc	femSc	
	845.19	856.28	mean
	119.97	85.15	std. dev.
	166	142	n
		296	df

Table 1: Test of Difference between Mean Scores			
Hypothesis Test: Independent Groups (t-test, unequal variance)			
		-11.09000	difference (maleSc - femSc)
		11.73728	standard error of difference
		0	hypothesized difference
		-0.94	t
		.3455	p-value (two-tailed)

The distributions of total online time use over the semester (in minutes) for pooled data, male and female students are shown in Figures 4, 5 and 6, respectively.

Figure 4: Distribution of Time Use (Pooled)

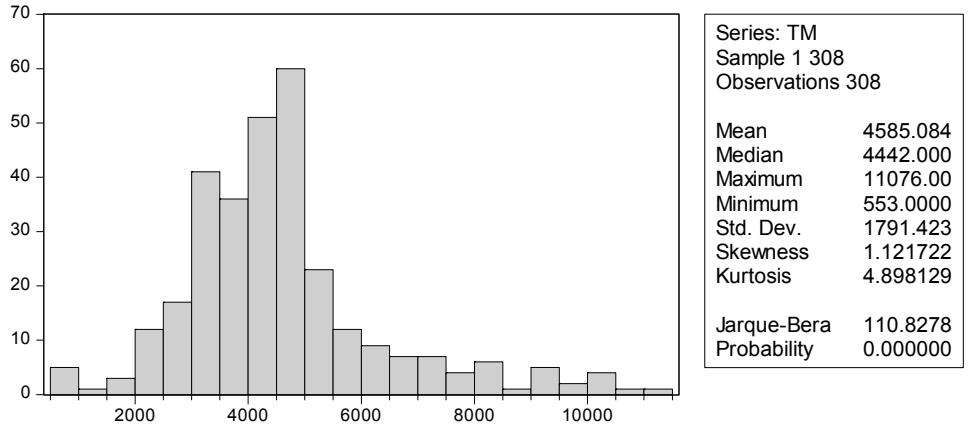


Figure 5: Distribution of Time Use (Male)

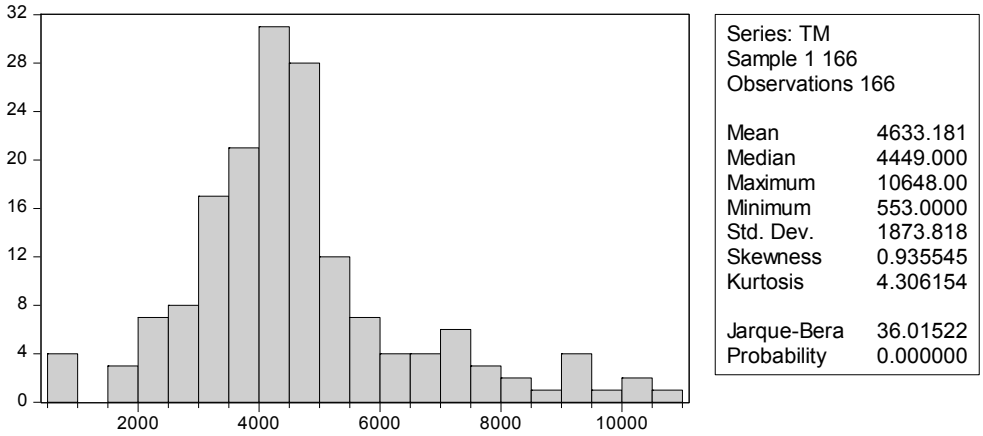
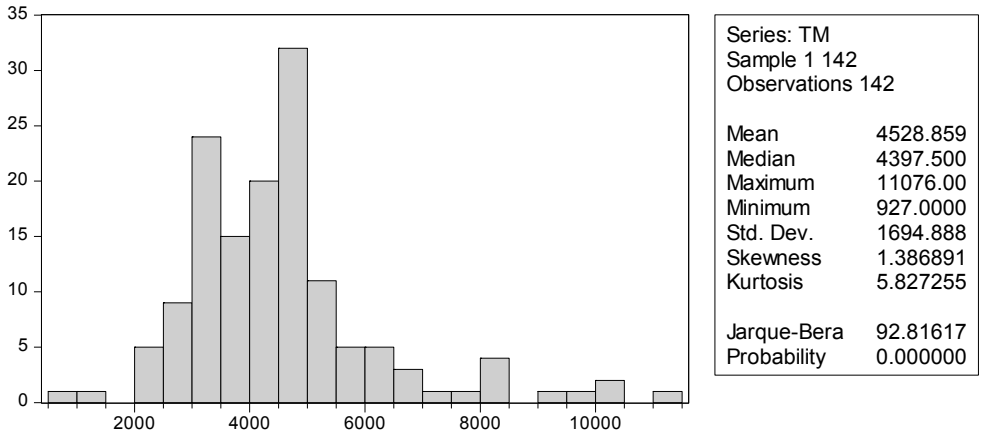


Figure 6: Distribution of Time Use (Female)



On average a female student spends about 4529 minutes online over the semester while a male student spends about 4633 minutes with the average for the two of 4585 minutes or 76.4 hours. The gender difference of 104 minutes over the semester is statistically quite insignificant as demonstrated in Table 2.

Table: 2 Test of Difference between Mean Time uses			
Hypothesis Test: Independent Groups (t-test, unequal variance)			
	maletm	femtm	
	4633.18	4528.86	mean
	1873.82	1694.89	std. dev.
	166	142	n
		305	df
		104.32000	difference (maletm – femtm)
		203.42507	standard error of difference
		0	hypothesized difference
		0.51	t
		.6084	p-value (two-tailed)

The correlation between score and time use for pooled data is reported in Table 3 which shows that the correlation is 0.59 and is highly significant statistically. The results for male and female students are very close and, therefore, are not reported here.

Table 3: Correlation between Online Time Use and Score		
Sample:		1 308
Included observations:		308
Correlation		
t-Statistic		
Probability	SC	TM
SC	1.000000	
TM	0.585637	1.000000
t-value	12.63855	-----
p-value	0.0000	-----

The Estimated Linear Regression Model

Table 4 reports the results of estimating the Linear Regression model.

Table 4: Linear Mixed Dummy Model				
Dependent Variable: SC				
Method: Least Squares				
Sample:				1 308
Included observations:				308
Variable	Coefficient	Std. Error	t-Statistic	Prob.
TM	0.037718	0.003541	10.65322	0.0000
GE*TM	-0.007683	0.005520	-1.392000	0.1649
GE	49.82585	27.05056	1.841952	0.0665
C	670.4318	17.68725	37.90481	0.0000
R-squared	0.351952	Mean dependent var		850.3019
Adjusted R-squared	0.345557	S.D. dependent var		105.3423
S.E. of regression	85.21943	Akaike info criterion		11.74124
Sum squared resid	2207755.	Schwarz criterion		11.78968
Log likelihood	-1804.151	Hannan-Quinn criter.		11.76061
F-statistic	55.03370	Durbin-Watson stat		2.008895
Prob(F-statistic)				0.000000

This model explains 35.2 % of variation and 34.6 % of variance in Score, The F- value shows that Coefficient of Determination R^2 is highly significant. The slope coefficient of time use is highly significant while the coefficient of gender (Intercept Dummy) is significant only with α at 10% level. The coefficient of Slope Dummy is insignificant. Thus, the impact of time use on score is highly significant but that of Gender is not significant according to this model.

We found, however, that the second model, which is more successful (as reported below), shows significant contribution of Gender.

Durbin Watson Statistic shows lack of the problem of auto-correlation. We have also performed Heteroscedasticity test using Harvey's method as reported in Table 5 below. The Harvey's test regresses logs of squared residuals on the original regressors. Harvey's test shows lack of the problem of Heteroscedasticity with respect to any regressor.

Table 5: Test of Heteroscedasticity

Heteroskedasticity Test: Harvey				
F-statistic	1.084411	Prob. F(3,304)		0.3559
Obs*R-squared	3.261141	Prob. Chi-Square(3)		0.3531
Scaled explained SS	3.124172	Prob. Chi-Square(3)		0.3729
Test Equation:				
Dependent Variable:		LRESID2		
Method:		Least Squares		
Date: 02/17/12		Time: 18:17		
Sample:				1 308
Included observations:				308
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.418806	0.450638	14.24382	0.0000
TM	0.000139	9.03E-05	1.534351	0.1260
GE*TM	-8.19E-05	0.000140	-0.584241	0.5595
GE	0.189423	0.682795	0.277423	0.7816
R-squared	0.010588	Mean dependent var		6.970303
Adjusted R-squared	0.000824	S.D. dependent var		2.177829
S.E. of regression	2.176931	Akaike info criterion		4.406611
Sum squared resid	1440.665	Schwarz criterion		4.455054
Log likelihood	-674.6182	Hannan-Quinn criter.		4.425981
F-statistic	1.084411	Durbin-Watson stat		1.980357
Prob(F-statistic)				0.355902

Interpretation of Model 1 results:

We have,

$$\hat{Sc} = 670.432 + 0.038Tm - 0.008Ge*Tm + 49.826Ge \qquad (4)$$

The slope coefficients indicate that an increase of 100 minutes in time use over the semester would cause predicted score to rise by 38 out of 1000 for a male student only 30 out of 1000 for female students. Since the coefficient of Slope Dummy is insignificant we cannot read much into this small gender difference in the expected reward for increased online time use. In order to examine the role of the intercept terms we will try to estimate the predicted score when a student has minimal time use. When online Time use over the semester is only about 553 minutes (which is the lowest value in the sample and is also nearly equal to the time taken by the Midterm and Final tests) the expected score for a male student is around 691 out of 1000 or 69.1%), and for a female student it is around 736/1000 showing a small difference of about 45 points out of 1000 (or 4.5%). However, as mentioned above, the contribution of Gender to intercept is significant only at 10% level. In the case of the second model, however, the conclusions about gender effect are quite strong.

The Estimated Constant Elasticity or Double-Log Regression Model

Table 6 reports the results of estimating the Log-Linear Regression model.

Table 6: Constant Elasticity (Double Log) Model

Dependent Variable				: LOG(SC)
Method:				Least Squares
Date: 02/17/12				Time: 18:11
Sample:				1 308
Included observations:				308
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(TM)	0.426926	0.027708	15.40781	0.0000
GE*LOG(TM)	-0.177090	0.046398	-3.816787	0.0002
GE	1.506889	0.387743	3.886311	0.0001
C	3.150760	0.231738	13.59620	0.0000
R-squared	0.482983	Mean dependent var		6.729786
Adjusted R-squared	0.477881	S.D. dependent var		0.223962
S.E. of regression	0.161830	Akaike info criterion		-0.791641
Sum squared resid	7.961422	Schwarz criterion		-0.743198
Log likelihood	125.9127	Hannan-Quinn criter.		-0.772271
F-statistic	94.66292	Durbin-Watson stat		1.971302
Prob (F-statistic)				0.000000

The model has much larger R^2 as well as $\bar{R}^2$ indicating that the model can explain 48.3% of variation and 47.8% of variance in log of Score. All the coefficients are highly significant

showing strong time use and gender effects on score. The F-statistic is quite high and the Akaike and Schwarz criteria are far better than the above model. The test for Heteroscedasticity was not carried out for this model as the regressor time is scaled down by taking log. The DW statistic shows lack of auto-correlation.

Interpretation of Model 2 results:

We have,

$\widehat{\log Sc} = 3.151 + 0.427\log(Tm) - 0.178Ge*\log(Tm) + 1.507Ge$ (5)

The double-log model provides some quite interesting results. First, the elasticity with respect to male student is 0.43 indicating that a 10% increase in Time use is expected to result in an increase of about 4.3 % of "existing" score (which will be slightly less than 4.3 percentage point because the "existing" score will be less than 100%). This is quite encouraging! Roughly speaking, a male student who is marginally (say 10%) below letter grade A could jump to A by making about 30% more efforts as measured by time use (given the assumption of this study that overall effort is proportionately related to online Time use). For a female student the prospect is good but a little dampened because of the negative Slope Dummy. For a female student the elasticity is 0.43 less 0.18 or 0.25. Thus a 10% increase in Time use is predicted to improve existing score by 2.5% as compared to 4.3% for male students. Moreover, the gender difference is also statistically quite significant. Thus, for example, an improvement of 10 percentage point in score would require an increase in Time use by nearly 50%. For an improved letter grade this is not a bad deal, provided the student has available time to use.

In order to see the use of the intercept we will try to estimate the predicted score when a student uses minimal online time use. If a male student spends only 553 minutes in online time use over the semester, the predicted score, using (natural) anti-log function comes to only about 346 out of 1000, or letter grade F. For a female student with similar time use the predicted score jumps to a little over 500, but is still letter grade F. These predictions are much lower than those of the above model.

The Estimated Decreasing Elasticity or Linear-Log Regression Model

Table 7 reports the results of estimating the Log-Linear Regression model.

Table 7: Decreasing Elasticity (Linear-Log) Model				
Dependent Variable:				SC
Method:				Least Squares
Date: 02/21/12				Time: 23:31
Sample:				1 308
Included observations:				308
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(TM)	209.9604	11.36220	18.47884	0.0000
GE*LOG(TM)	-37.19491	19.08718	-1.948685	0.0523
GE	321.5180	159.6213	2.014254	0.0449
C	-908.4865	95.04119	-9.558871	0.0000
R-squared	0.607490	Mean dependent var		850.3019
Adjusted R-squared	0.603616	S.D. dependent var		105.3423
S.E. of regression	66.32243	Akaike info criterion		11.23984
Sum squared resid	1337194.	Schwarz criterion		11.28828
Log likelihood	-1726.935	Hannan-Quinn criter.		11.25921
F-statistic	156.8341	Durbin-Watson stat		1.942221
Prob(F-statistic)				0.000000

The Decreasing Elasticity model has the largest R^2 as well as $\bar{R}^2$ indicating that it can explain 60.7% of variation and 60.3% of variance in Score which is quite high for a cross-sectional study where numerous individual characteristics are at play but only a couple are measured and used. The F-statistic exhibits high statistical significance of the coefficient of Determination. The coefficient of Log(Time) is highly significant , but the coefficients of other regressors have lost some statistical significance compared to the previous (double-log) model. Now, the Slope Dummy coefficient is significant only at 10% while Intercept Dummy is significant at 5% levels. Similarly, the Akaike and Schwarz criteria have become high (or worse) compared to the previous model. The DW statistic is still near 2. The Heteroscedasticity test for this model too was not performed for the reason mentioned above.

Interpretation of Model 3 results:

We have,

$$\hat{Sc} = -908.487 + 209.960\log(Tm) - 37.195Ge*\log(Tm) + 321.518Ge \quad (6)$$

The elasticity with respect to log(Tm) for a male student is 209.960/Sc, which continuously decreases as the level of score increases. For a minimal level of Time use of 553

minutes over the semester the calculated elasticity 209.960/553 or equal to 0.38, indicating that such a male student is expected to improve the score by 3.8% of existing score with additional 10% online time use. Plugging the values, the predicted score for a male student with minimal online time use is only 417/1000 which is only 7.1 percentage point above the previous model but is still letter grade F. On the other hand, a male student with time use around the overall mean of 4585 minutes will have elasticity only 209.96/4585 equal to 0.05 indicating that a further 10% increase in Time use is predicted to improve the score by only 0.5% of existing score. The predicted score for such a student is midway between letter grade B and letter grade A. For a student with even higher level of time use, the reward for additional online time use will be obviously pretty low.

Similarly, for a female student with online time use of only 553 minutes the elasticity comes to (209.96- 37.195)/553 equal to 0.31 indicating that a 10% increase in time use is expected to improve the existing score by 3.1% of existing score. The predicted score for such a female student is, however, 86.6/1000 above a similar male student or equal to 504/1000 which is pretty close to the prediction of the previous model. Thus a female student too, with minimal online time use can significantly improve her score with additional efforts. Again the elasticity continuously declines as the online time use level increases, with values close to those of the male students. In fact, the gender difference in elasticity also declines as the level of score increases.

CONCLUSIONS

This study uses actually recorded online time use of students instead of self-reported surveys used in most studies in the relevant literature. Moreover, the models use actual scores instead of the letter grades which not only hide a lot of information by converting the ratio scale variable to discrete ordinal variable. As a result, this study could safely use various forms of Multiple Regression models. A basic assumption underlying this study is that online time use is objectively measurable and good indicator of overall effort by students in online classes. The evidences suggest that there is a significant reward for additional effort, especially at the lower levels of times use and scores. The Constant Elasticity model predicts a 4.3% improvement in existing score for additional 10% increase in online time use for male students. For female students the improvement is expected to be only about 2.5% in existing score. The gender difference is highly significant statistically in the Constant Elasticity model. The decreasing Elasticity model is not only theoretically more appealing but also most successful in explaining variations in the scores. It can explain about 60% of variation in scores, which is quite high for a cross-sectional study where numerous individual characteristics are at play while only a couple of attributes are measured and used.

However, the gender difference gets dampened and loses some of its statistical significance in the Decreasing Elasticity model compared to the Constant Elasticity model.

According to the decreasing Elasticity model, a 10% increase in online time use for male students with minimal online time use, is expected to improve the existing score by 3.8% of existing score. For a similar female student the predicted improvement is 3.1% of existing score. As the level of time use increases to the mean level (4585 minutes over the semester or 76. 4 hours), the elasticity for male students drops to 0.05 indicating that a 10% increase in time use would be expected to improve existing score only by 0.5%. The gender difference becomes very small at higher levels of time use. The results of this study are particularly significant for students with low online time use. Instructors should encourage such students to significantly increase their effort as it promises much larger reward. Although few students can and have achieved high scores despite their low online time use, it is clear from the data that very low online time use is a good predictor of low scores with few exceptions. This study can be easily extended by incorporating other objectively measurable attributes of the students, such as their previous GPA, Race, Level of Education (Graduate vs. Undergraduate) and also covering other subjects.

REFERENCES

Asteriou, D. and S. G. Hall (2007). Applied Econometrics: A Modern Approach. Palgrave-Macmillan, 161-65

Brint, S. and A. Cantwell (2010). Undergraduate Time Use and Academic Outcomes: Results from UCUES 2006 *Teachers College Record*, 112(9), 2441-2470

Brint, S., et al. (2010). Engaged Learning in a Public University: Trends in the Undergraduate Experience. Report on the Results of the 2008 University of California Undergraduate Experience Study. Center for Studies in Higher Education, Report.

Brookshire, R. G. and, S. W. Palocsay (2005). Factors Contributing to the Success of Undergraduate Business Students in Management Science Courses. *Decision Sciences Journal of Innovative Education,* 3, 99-108.

Buckles, S. G. and M.E. McMahon (1971). Further Evidence on the Value of Lecture in Elementary Economics. *Journal of Economic Education*, 2(2), 138-41.

Burrus, Jr., R. T., C. F. Duma and J. E. Graham, Jr. (2001). *The Journal of Economics and Economic Education Research, 1-10*

Cheo, R. (2003). Making the Grade through Class Effort Alone. *Economic Papers,* 22(2), 55-65.

Cheung, L., & Kan , A. (2002). Evaluation of factors related to student performance in a distance-learning business communication course. *Journal of Education for Business, 77*(5), 257–263.

Chan, Kam C., C. Shum, and D. J. Wright (1997). Class Attendance and Student Performance in Principles of Finance" *Financial Practice and Education*, 7, 58-65.

Cohn, E. and E. Johnson (2006). Class Attendance and Performance in Principles of Economics

Credé, M., S. G. Roch, and U. M. Kieszczynka (2010). Class Attendance in College: A Meta Analytic Review of the Relationship of Class Attendance with Grades and Student Characteristics

Damianov, D. S, et al. (2009). Time Spent Online and Student Performance in Online Business Courses: A Multinomial Logit Analysis. *Journal Of Economics And Finance Education,* 8(2), 11-22.

Day, S., 1994. Learning in large sociology classes: Journals and attendance. *Teaching Sociology*, 22 (2), 151-165.

Devadoss, S., and J. Foltz. (1996). Evaluation of Factors Influencing Student Class Attendance and Performance. *American Journal of Agricultural Economics* 78: 499-507.

Douglas, S. and J. Sulock (1995). Estimating educational production functions with correction for drops. *Journal of Economic Education*, 26 (2), pp.101-112.

Douglas, I. and N.D. Alemanne (2007) IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA), 299-302

Durden, G. C. and L. V. Ellis (1995). The Effects of Attendance on Student Learning in Principles of Economics. *American Economic Review Papers and Proceedings*, 85(2), 343-346.

Ellis, L. V., G. C. Durden and P. E. Gaynor (1998). Evidence on Factors That Influence the Probability of a Good Performance in the Principles of Economics Course. *Business Quest*

Ellis, L. V. and G.C. Durden (2002). Improving student attendance: does it improve student learning? *Academic Exchange Quarterly*

George, D., et al. (2008). Time Diary and Questionnaire Assessment of Factors Associated with Academic and Personal Success among University Undergraduates. *Journal of American College Health*, 56(6), 706-715

Gatherer, D. and F. C. R. Manning (1998). Correlation of examination performance with lecture attendance: a comparative study of first-year biological sciences undergraduates. *Biochemical Education*, 26 (2), 121-123

Gump, S. E. (2005). The cost of cutting class: attendance as a predictor of student success. *College Teaching,* 53, 21–26.

Gunn, K. P. (1993). A correlation between attendance and grades in a first-year psychology class. *Canadian Psychology,* 34(2), 201-202.

Johnson, D. L., P. B. Joyce, and S. Sen (2002). An Analysis of Student Effort and Performance in the Finance Principles Course. *Journal of Applied Finance: Theory, Practice, Education*, 12(2), 67-72

Launius, M.H. (1997). College student attendance: attitudes and academic performance. *College Student Journal*, 31 (1), .86-92

Ledman, R. E. and F. Kamuche (2002). Improving Student Attendance: Does it Improve Student Learning? *Academic Exchange Quarterly*, 6(1) 76-80

Lin, T. F. and J. Chen (2006). Cumulative Class Attendance and Exam Performance. *Applied Economics Letters,* 13, 937-942.

Marburger, D. R. (2001). Absenteeism and Undergraduate Exam Performance. *Journal of Economic Education*, 99-109.

Marburger, D. R. (2006). Does Mandatory Attendance Improve Student Performance? *Journal of Economic Education*, 37(2), 148-155

Moore, R. (2003). Attendance and performance: How important is it for students to attend class? *Journal of College Science Teaching, 32,* 367–371.

Moore, R, et al. (2003). Showing up: The importance of class attendance for academic success in introductory science courses. *American Biology Teacher, 65*, 325–329.

Nonis, S. A., et al. (2005). The Impact of Non-Intellectual Variables on the Academic Success of Business Students. *Marketing Education Review,* 15, 51-63.

Nonis, S. A., M. J. Philhours, and G. I. Hudson. 2006. A diary approach to business and marketing students' time use. *Journal of Marketing Education* 28:121-134.

Nyamapfene, A. (2010). Engineering Education. *Journal of the Higher Education Academy,* Engineering Subject Centre, 5(1)

Park, K. H. and P. M. Kerr (1990). Determinants of Academic Performance: A Multinomial Logit Approach. *Journal of Economic Education*, 21(2), 101-11.

Picciano, A.G. (2002). Beyond student perceptions: issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Networks*, 6 (1), 21-40

Rich, S. P. (2006). Student Performance: Does Effort Matter? *Journal of Applied Finance* 16, 120-133.

Riffell, S. and D. Sibley (2005). Using web-based instruction to improve large undergraduate biology courses: an evaluation of a hybrid course format. *Computers & Education*, 44 (3), 217-235

Rodgers, J.R. (2001). A panel-data study of the effect of student attendance on university performance. *Australian Journal of Education*, 45 (3), 284-295

Rodgers, J.R. (2002). Encouraging tutorial attendance at university did not increase performance, *Australian Economic Papers*, 41 (3), 255-266.

Romer, D. (1993). Do Students Go to Class? Should They? *Journal of Economic Perspectives*,7(2), 167-174.

Schmidt, R.M. (1983). Who Maximizes What? A Study in Student Time Allocation. *American Economic Review*, 73(2), 23-28.

Stinebrickner, R. and T. R. Stinebrickner. (2004). Time-Use and College Outcomes. *Journal of Econometrics*, 121(1-2): 243-269.

Stinebrickner R. and T.R. Stinebrickner. (2008). The Causal Effect of Studying on Academic Performance. *The B.E. Journal of Economic Analysis and Policy (Frontiers)* 8(1), 14: 1-53.

A STUDENT FRIENDLY ILLUSTRATION AND PROJECT: EMPIRICAL TESTING OF THE COBB-DOUGLAS PRODUCTION FUNCTION USING MAJOR LEAGUE BASEBALL

James B. Larriviere, Spring Hill College
Ralph Sandler, Spring Hill College

ABSTRACT

There has been a plethora of thinking and research about better methods of teaching important economic concepts to undergraduate and graduate level students. One concept students have a particularly difficult time grasping is the Cobb-Douglas (C-D) production function. Not only does the concept create a hurdle for students, but the actual estimation and interpretation of the C-D production function has perplexed even the most dedicated students.

This paper offers a pedagogical tool to make the instruction and estimation of (C-D) production functions much more palatable to students in the intermediate level undergraduate economic courses and graduate students in managerial economics. To that end, the paper has two parts. In the first part, we introduce the production function and make the connection between team sports and the productivity of inputs. Next, we relate a specific type of production function, the Cobb-Douglas, to major league baseball teams' performance. A model is introduced and explained, with a description of the various inputs used to produce team wins. The first part concludes with the estimation of a log-linear C-D production function and an evaluation of the results.

The paper's pedagogical contribution is explained in the second part of the paper. We provide a student project that professors can use in their courses to facilitate the teaching of C-D production functions. The project comes complete with data from the 2006 MLB season, which students use to estimate a log-linear C-D production function. The steps involved in the data transformation process are provided along with the empirical results. Questions and answers about the regression results are included and could be used to reinforce students' learning and understanding of the model and output.

INTRODUCTION

Cobb-Douglas Production Function

Considering that most students have played a team sport at some point, this familiarity will make it easier for the student to grasp the connection between what the manager/coach is trying to produce (team wins) and productive inputs (which are measured by players' offensive and defensive skills). As with most team sports, baseball being no exception, team composition

is critical for maximizing wins, so the manager's task is to obtain players with the right mix of skills and abilities, bring them together, and win ballgames. For the manager, understanding the impact of inputs such as hitting, pitching, errors and stolen bases on a team's winning percentage is important for at least two reasons. First, if empirical results indicate that pitching has a greater positive impact on team wins than hitting, then the manager can trade or hire accordingly. For example, a manager may trade a player with a strong batting average in order to get a pitcher with a really low earned run average. Second, those players whose skills have the greatest positive impact on team wins (a high marginal product) may negotiate for higher salaries, which would certainly affect the team payroll. Managers (or owners) do not want to overpay for a player if his skills will not have much impact on team wins.

Let us begin with a brief explanation of the production function and how it can be used to examine major league baseball team success. The production function shows the resulting quantity of output that can be produced with a given set of inputs. It also provides a conceptual framework for decisions involving the allocation of the firm's (team's) resources. One particular form of the production function is the Cobb-Douglas (C-D). The C-D is used extensively in economic studies, and its properties have kept it popular for over eighty years.

Given the importance of production theory to microeconomic analysis and the obvious practical appeal of using the Cobb-Douglas, it is interesting to note the range of coverage found in a sample of economic textbooks. The Pindyck and Rubinfeld text (2005) is typical of the very brief discussion found in undergraduate intermediate microeconomic books. MBA level managerial economic textbooks tend to provide more coverage, sometimes including the results of empirical research using the C-D (see for example, Thomas and Maurice (2005); McGuigan, Moger and Harris (2006); Hirschey (2003); Baye (2009); Keat and Young (2006)). The textbooks listed above do not have an extensive case-like project that requires students to take a set of actual data, which is provided to them, and empirically estimate a C-D production function.

Economists have used the C-D to estimate the impact various inputs have on major league baseball teams' success. Two noteworthy articles by Charles Zech (1981) and Mark Woolway (1997) use cross-sectional data from the 1977 and 1993 baseball seasons, respectively, to empirically estimate a C-D production function. Both articles sought to establish a relationship between team victories (output) and various inputs, like hitting, pitching, and defensive skills. The variables used in our model will follow Zech (1981) and Woolway (1997) model specifications.

The C-D is typically illustrated by the following equation:

$$Q = AL^bK^c$$

where Q is output, A is the total factor productivity (the change in output not caused by the inputs e.g. by technology change or weather), and L and K are inputs, typically labor (L) and capital (K). The exponents, b and c, are to be estimated. The C-D assumes some degree of

substitutability between the inputs, albeit not perfect substitutability one finds in a strictly linear model specification. Since the C-D is a multiplicative model, not a linear model, taking the logarithms of the data is necessary in order to estimate the function using OLS linear regression. The standard log-linear model is:

$$\text{Log } Q = \log A + b \log L + c \log K.$$

MODEL

To estimate major league baseball's production function, data was obtained from *The ESPN Baseball Encyclopedia (2007).* The output measure (Q) is team victories in the regular season. The input measures consist of four primary categories: offense, pitching, defense and managerial effectiveness.

Offense requires hitting for average and hitting for power; consequently, we use team batting average (BA) and home runs (HR). We also use stolen bases (SB), an offensive sub-skill that one would expect to have an influence on winning games. We expect all of these to have a positive relationship to team victories.

Earned run average (ERA) is considered to be the most important measure of pitching effectiveness (Woolway 1997). ERA is measured by dividing the number of earned runs the pitcher allows by the number of innings pitched and is then multiplied by nine (the regulation length of a Major League Baseball game). The lower the team ERA the better; therefore, the variable should have an inverse relationship to winning percentage. Defense, the third major input category, is measured by the total number of errors committed by the team (E) and should similarly have an inverse relationship to winning percentage.

Zech (1981) did not find managerial effectiveness to be a significant factor in team wins. Nevertheless, we think some specification for this input should be considered and have included the manager's lifetime winning percentage (MW) in the equation. The variable should have a positive relationship to team victories.

In order to control for the fact that variables are measured in different units, we have indexed them by dividing all team variables (dependent and independent) by either their National or American League averages.

The Cobb-Douglas production function for Major League Baseball will be defined as follows:

$$V = A * BA^{a} * HR^{b} * SB^{c} * ERA^{d} * E^{e} * MW^{f} \qquad (1)$$

In order to estimate the parameters of this nonlinear equation, it will be transformed into a linear form. This is accomplished by taking the natural logarithms (log) of both sides of equation (1). The transformed equation is:

$$\text{Log } V = \log A + a \log BA + b \log HR + c \log SB + d \log ERA + e \log E + f \log MW \quad (2)$$

where

V = team victories/league average
BA = batting average/league average
HR = home runs/league average
SB = stolen bases/league average
ERA = earned run average/league average
E = errors/league average
MW = manager's lifetime winning percent/league average

EMPIRICAL RESULTS

An OLS regression was performed on equation (2), and the estimated coefficients for a, b, c, d, e, and f are provided in Table 1. Three key variables, batting average (BA), home runs (HR) and pitching (ERA) were found to be statistically significant (at the .05 level). The variables had the correct sign, except for stolen bases.

The relative size of the estimated parameters is also revealing. Consider the estimate for batting average (BA). A 10% increase in team batting average is associated with a 14.27% increase in team winning percentage. As for the impact of home runs on team winning percentage, a 10% increase in home runs is associated with a 2.07% increase in team winning percentage. The impact of hitting for average (BA) is nearly seven times greater than hitting for power (HR). The long lasting debate among baseball fans as to the relative value of hitting and pitching is also addressed. The two hitting variables combined (BA and HR) have a greater impact than pitching, a result consistent with the findings of both Zech (1981) and Woolway (1997). It is also interesting to note that three input variables (SB, E, and MW%) were not found to have a significant effect on team winning percentage, which is consistent with Zech's (1981) earlier findings.

Our results so far have established a relationship between team victories (output) and various inputs for just one year (2005). However, further evidence is necessary before coming to any strong conclusions about the keys to success for major league baseball teams. Our statistical evidence should be reaffirmed over several years. Therefore, the same cross-sectional regression model is applied using data for the 2000, 1996, 1990 and 1985 seasons. [1] Results are provided in Table II, a summary of which follows:

1. The results of our regression analysis using the same variables over four additional years are rather consistent. The R^2 for all five years ranges from .82 to .91. Batting average

(BA), home runs (HR), and earned run average (ERA) all have the correct sign and are statistically significant at the .05 level in each of the five equations.

2. The independent variable stolen bases (SB) was only significant in one year (1996), while the number of errors committed by a team (E) was significant in two years (1996 and 1985). In each case, the variable reflected the correct sign.
3. The independent variable MW, which represents the manager's lifetime winning percent, was only significant in the 1996 season.

In theory, the role of the manager should be indispensable to a team's success. Managers have the responsibility of training and motivating players, maintaining player morale, as well as making tactical decisions during a game to enhance the team' success. Yet, we do not find strong statistical evidence in our results that supports the theory. It may well be managers serve only a marginal role in a team's success. However, it seems unreasonable to believe that rational, profit-maximizing owners would pay relatively large salaries to managers despite their questionable value to the organization. Another possible explanation is that our results may just reflect the variable used as a proxy for managerial efficiency (MW). The absence of solid quantitative data makes it very difficult to come up with alternative measures. Finding an effective measure of managerial efficiency would be an excellent topic for further research.

STUDENT PROJECT

Estimating log-linear Cobb Douglas Production Function

This second part of the paper describes a project that requires students to estimate a Cobb-Douglas production function using 2006 MLB season data. By having them estimate and evaluate a regression equation using real economic data, students are much more likely to understand the economic and statistical concepts.

It is suggested that instructors provide students with copies of **Part I** of this paper. The instructor may assign the Instroduction to the students, discuss the paper in class, and then assign the student project described here in the Student Project**.** The raw data for the 2006 baseball season is provided in Appendix A. *Excel,* as well as other spreadsheet statistical packages, can be used to transform the data and run the statistical analysis.

Data Transformation

The data in Appendix A are measured in different units; therefore, it is customary to use a type of index in order to utilize the data in the model. The first step toward indexing the variable values is to find the league average for each variable. Next, each team's variable values are divided by the league average. For example, if Boston's wins 86 games, and the American

League's average number of victories is 83, then Boston's indexed number of wins is 86/83 = 1.036. This process is applied to all seven variables.

The second step in the data transformation process is determining the natural log of the data. The natural log is determined for each indexed variable. In the *Excel* package, the LN function is applied. Once all data are transformed into their log form, regression analysis can be performed.[2]

Regression Results and Sample Questions

Regression results from *Excel* are provided in Appendix B for instructors. The instructor may use the following questions to evaluate students' understanding of the model and interpretation of the estimated coefficients (reasonable answers are given in italics after each question.)

Question 1: Calculate the direction and statistical significance of the independent variables.

In evaluating the significance of the independent variables a useful "rule of thumb" is that if the absolute value of the t-statistic is greater or equal to 2, then the parameter estimate is statistically different from zero at the .05 level of significance (Baye, p. 100). In addition, a low P-value (lower than .05) suggests only a small chance that the true coefficients are actually zero. By these standards, the independent variables HR and ERA are statistically significant, where as the variables BA, SB, MW are not. The variable E is significant at the .10 level of significance. As hypothesized above, the coefficients of all variables have the correct sign, indicating that the model is consistent with our theory about the Cobb-Douglas production function as applied to Major League Baseball.

Question 2: Evaluate the overall performance of the model.

The R-square (coefficient of determination) and F-statistic tell us about the overall performance of the model. The R-square, which tells us the fraction of the total variation in the dependent variable explained by the regression, is .80 in Appendix B. In addition, the F-statistic, which allows one to objectively determine the statistical significance of any regression, suggests there is an infinitely small chance (3.97E-07) that the estimated regression model fits the data purely by accident.

Question 3: Define output elasticity. Given the regression results for 2006 (Appendix B), what impact would a 10 percent increase in HR have on output?

In a Cobb-Douglas production function, where the data are transformed by taking the natural log of all variables, the coefficients are all output elasticities. An output elasticity measures the percentage change in output (team victories) divided by the percentage change in some input variable. For example, the output elasticity for the coefficient "b" (HR) is

$$b = \frac{\text{percentage change in output (victories)}}{\text{percentage change in home runs}}$$

So that a 10 percent increase in a team's HR, ceteris paribus, should lead to a 2.29% increase in output where

Percentage change in output = (percentage change in HR) x b

= (+) 10% x .229
= 2.29 percent

Question 4: Reflect on your results within the broader context of empirical research. Specifically, compare and contrast your results with the five regression equations discussed in this paper.

As in the other five equations, the Cobb-Douglas production function using the 2006 Major League Baseball season provides a reasonably good fit (R^2=.80, low significance value of the F-statistic). In addition, HR and ERA are statistically significant (at the .05 level of confidence) and have the correct sign. However, unlike the other years, batting average (BA) is not significant in this equation.

CONCLUSION

This paper has two purposes. The first is to introduce the Cobb-Douglas production function to students in a team sport application: major league baseball. By using the popular national pastime as an illustration, we think students will become more interested in the broader topic of production analysis. In estimating the Cobb-Douglas production function, cross-sectional regressions were run for five different years (1985, 1990, 1996, 2000, and 2005) thus strengthening the confidence we have in our results that are reasonably consistent over time. Our empirical results demonstrate that:

1) Batting average (BA), home runs (HR) and earned run average (ERA) are statistically significant in each of the five years examined.
2) Consistent with previous research, the relative size of the coefficients suggests hitting (both BA and HR) contributes more to team victories than pitching.

3) In four of the five years examined, our measure of managerial effectiveness was not significant. We attribute this to the proxy used (MW). Finding an alternative measure of managerial effectiveness would be an excellent topic for further research.

In the second part of the paper we propose a student project that utilizes the Cobb-Douglas specification with data from the 2006 Major League Baseball regular season. Based on the model previously specified, students are provided the raw data and then required to transform the data, estimate the model and ultimately explain the economic and statistical concepts. We believe such a project will enhance their understanding of the material beyond that achieved by a lecture alone. The hands-on exercise centered around a popular sport is more likely to grab the attention of students than more traditional economics queries. Additionally, being able to compare the findings of past research with regression coefficients that the students themselves estimate is an effective teaching tool.

ENDNOTES

[1] Although our initial strategy was to select baseball seasons at five-year intervals, the 1995 season was shortened by the players' strike that may have given us abnormal results. Consequently, we used the 1996 season instead.

[2] Within *Excel*, have students click **tools**, and then **data analysis** (some computers may require operators to add-in the analysis tool pak) followed by **regression**. The dependent variable column should be added to the input Y-range and all independent variable columns should be placed in the input X-range. It is suggested that they request **labels** and a **95% confidence interval**.

REFERENCES

Baye, M. 2009. *Managerial Economics and Business Strategy.* 6th edition. Boston, MA: McGraw Hill Irwin.

Cobb, Charles W. and Douglas, Paul H. 1928. Theory of Production. *American Economic Review* 8:139-165.

Hirschey, Mark. 2003. *Managerial Economics*. Mason, OH. Thomson-Southwestern.

Keat, Paul G. and Young, Philip. 2006. *Managerial Economics, Economic Tools for Today's Decision Makers.* 5th edition. Upper Saddle River, NJ. Pearson- Prentice Hall.

McGuigan, J., Moyer, R., Harris, F. 2005. *Managerial Economics, Applications, Strategy And Tactics.* 10th edition. Mason, OH. Thomson-Southwestern.

Pindyck, R. and Rubinfeld, D. 2005. *Microeconomics.* 6th edition. Upper Saddle River, NJ. Pearson-Prentice Hall.

Thomas, C. and Maurice, S. 2005. *Managerial Economics.* 8th edition. Boston, MA. McGraw Hill-Irwin.

The ESPN Baseball Encyclopedia. 2007. 4th edition, edited by Gary Gillett and Pete Palmer. New York, NY. Sterling Publishing Co.

Woolway, Mark D. 1997. Using an Empirically Estimated Production Function for Major League Baseball to Examine Worker Disincentives Associated with Multi-Year Contracts. *The American Economist.* 41:77-83.

Zech, Charles E. 1981. An Empirical Estimation of a Production Function. *The American Economist.* 25:19-23.

Table 1: Regression Results **2005**			
Parameter	**Estimate**	**t-ratio**	**P-value**
A (Intercept)	-0.09	-0.686	0.500
a (BA)	1.427	4.316*	0.001
b (HR)	0.207	3.686*	0.001
c (SB)	-0.008	0.365	0.72
d (ERA)	-0.671	-6.385*	2E-06
e (E)	-0.224	-1.778	0.089
f (MW)	0.237	1.647	0.113
F-value 24.60 R square .87 *Significant at .05 level			

Table II: Regression Results **Multiple Years**					
	2005	**2000**	**1996**	**1990**	**1985**
Parameters					
A (Intercept)	-0.009	-0.008	-0.005	-0.008	-0.012
a (BA)	1.427*	1.550*	1.500*	1.291*	1.741*
b (HR)	0.207*	0.183*	0.224*	0.163*	0.306*
c (SB)	-0.008	0.001	0.094*	-0.0001	0.036
d (ERA)	-0.671*	-1.227*	-0.690*	-0.891*	-0.916*
e (E)	-0.223	-0.008	-0.200*	-0.110	-0.491*
f (MW)	0.237	-0.166	0.347*	-0.013	-0.219
F-value	24.60	37.09	39.54	14.20	25.29
R-square	0.87	0.91	0.91	0.82	0.89

Appendix A: Raw Data 2006							
Team AL	**V**	**BA**	**HR**	**SB**	**ERA**	**E**	**MAN W/L%**
NY	97	0.285	210	139	4.41	104	0.536
BOS	86	0.269	192	51	4.83	66	0.497
TOR	87	0.284	199	65	4.37	99	0.514
BAL	70	0.277	164	121	5.35	102	0.429
TB	61	0.255	190	134	4.96	116	0.419
CHI	90	0.28	236	93	4.61	90	0.56
CLE	78	0.28	196	55	4.41	118	0.492
DET	95	0.274	203	60	3.84	106	0.493
KC	62	0.271	124	65	5.65	98	0.416
MIN	96	0.287	143	101	3.95	84	0.562
OAK	93	0.26	175	61	4.21	84	0.568
SEA	78	0.272	172	106	4.6	88	0.501
ANA	89	0.274	159	148	4.04	124	0.537
TEX	80	0.278	183	53	4.6	98	0.514
Team NL							
ATL	79	0.27	222	52	4.6	99	0.563
NY	97	0.264	200	146	4.14	104	0.556
FLA	78	0.264	182	110	4.37	126	0.481
WAS	71	0.262	164	123	5.03	131	0.475
PHIL	85	0.267	216	92	4.6	104	0.535
STL	83	0.269	184	59	4.54	98	0.536
CIN	80	0.257	217	124	4.51	128	0.473
MIL	75	0.258	180	71	4.82	117	0.45
HOU	82	0.255	174	79	4.08	80	0.483
PIT	67	0.263	141	68	4.52	104	0.508
CHI	66	0.268	166	121	4.74	106	0.527
SF	76	0.259	163	58	4.63	91	0.503
LA	88	0.276	153	128	4.23	115	0.568
ARI	76	0.267	160	76	4.48	104	0.477
COL	76	0.27	157	85	4.66	91	0.447
SD	88	0.263	161	123	3.87	92	0.494

Appendix B: Regression Results 2006					
Regression Statistics					
Multiple R	0.897012198				
R Square	0.804630883				
Adjusted R Square	0.753665026				
Standard Error	0.062587744				
Observations	30				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance of F*
Regression	6	0.371062625	0.061844	15.7876456	3.97278E-07
Residual	23	0.090096191	0.003917		
Total	29	0.461158816			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	
A (Intercept)	-0.007277222	0.011697174	-0.62214	0.53997215	
BA	0.799905128	0.494106737	1.618891	0.11910328	
HR	0.229302108	0.085878985	2.670061	0.01367548	
SB	0.040020952	0.036791781	1.087769	0.28796517	
ERA	-0.799607547	0.165322329	-4.83666	6.9971E-05	
E	-0.183639004	0.09186517	-1.99901	0.05755922	
MAN W/L%	0.280892048	0.19198527	1.463092	0.15697493	

THE EFFECTS OF TEACHER MERIT PAY ON ACADEMIC ATTAINMENT: AN ANALYSIS USING DISTRICT-LEVEL DATA

Mark Gius, Quinnipiac University

ABSTRACT

This study uses district-level data for the academic year 2007-2008 in order to determine if teacher merit pay has any effect on student graduation rates and drop-out rates. Using data from the Schools and Staffing Survey (SASS) and the American Community Surveys, the results of this study suggest that merit pay is not positively related to student academic attainment. Factors that have an effect on academic attainment include region of residence, racial composition of student body, educational attainment of district residents, and median family income. This study is important because it is the first to use a national set of district-level data, and its findings are consistent with some of the research in this area in that it finds that merit pay has no positive and significant effects on academic attainment.

INTRODUCTION

The vast majority of public school teachers in the US are compensated according to a single salary schedule. With this type of compensation system, a teacher's pay is based upon only two factors: years of experience and level of education; the quality of a teacher's instruction has no bearing on a teacher's compensation. Some believe that such a salary schedule does not promote individual achievement or excellence (Figlio and Kenny, 2007). Rather, they believe that a compensation system based more on teacher performance and less on seniority would do much to improve public education in the US. Under such a merit pay system, teachers would be monetarily rewarded if their students excel according to some measure of academic achievement. If, however, a teacher's students did not attain some minimum level of academic success, then that teacher would not receive any merit pay. If a teacher's students continually failed to meet some measure of academic success, then that teacher would slowly see their compensation decline in real terms over time. Hence, under a merit pay system, the high-performing teachers would see gains in their incomes over time while low-performing teachers would see their real incomes fall. The goal of such a system would be to give the low-performing teachers a reason to leave the profession. If these types of teachers left, then it is hoped that students would benefit from having mostly high-performing teachers in the classroom. Under the current single salary schedule system, low-performing teachers have much less of an incentive to quit the profession.

Thus, students suffer under such a system because of the greater prevalence of low-performing teachers.

The purpose of this study is to examine the effects of a merit-based compensation system for teachers on two measure of student academic attainment. The next two sections of this paper will provide further background on faculty merit pay and will discuss prior research on this topic. The empirical model will then be presented, and finally, the results will be discussed.

BACKGROUND INFORMATION

As noted in the introduction, under a merit pay system, a teacher's pay would be based on their teaching effectiveness. Low-performing teachers would receive little or no annual pay raises. Over a number of years, these teachers would see their salaries decline in real terms. Such teachers may eventually realize that their services are no longer desired and would leave the teaching profession. On the other hand, effective teachers would receive potentially substantial pay raises, which would encourage them to improve their teaching even more and would provide an incentive for others to enter the teaching profession. Given that the rewards for excellent teachers would increase under a merit pay system, it is reasonable to assume that highly qualified individuals would be more inclined to enter the teaching profession. Hence, under a merit pay system, the good teachers would be retained, the bad teachers would leave, and many potentially excellent teachers would be enticed to enter the teaching profession.

In addition to ridding the public schools of poor teachers, a merit-based compensation system would also, in all likelihood, reduce the instructional expenditures of local school districts. Instead of giving all teachers the same percentage pay increase every year, only a few teachers would receive substantial pay raises. Many others would receive much smaller raises, and some would receive no pay increases at all. Merit pay, in combination with other compensation reform measures currently being debated at the state level (elimination of collective bargaining and revamping of the state's teacher pension systems, for example) should reduce the overall cost of public education at the state and local level, thus helping to reduce the large deficits many state and local governments are currently facing.

Many teachers, however, oppose merit pay primarily because they believe that it would undermine one of the most important aspects of teaching: collaboration. Teachers would not be willing to collaborate under such a system because it would detract from their individual attainment and thus potentially reduce their compensation. In addition, if there is a fixed pool of funds for merit pay, then assisting a fellow teacher may result in less merit pay for everybody else. Of course, it may be possible to devise a merit pay plan that rewards collaboration, but whenever output is jointly produced, as education is, it is difficult to distinguish between individual levels of attainment.

There is research that supports the teachers' argument against merit pay and suggests that this type of compensation system is not a viable option in education (Murname and Cohen,

1986). Murname and Cohen (1986) claim that many factors affect student achievement and to hold teachers directly accountable for the academic success of their students is unfair and neglects the roles that others (other teachers, parents, school administrators, and the students themselves) have in achieving scholastic success. Second, as noted earlier, collaboration is very important in teaching. Teachers must be encouraged to collaborate. However, under most merit pay plans, the role of collaboration or joint production is ignored. A successful and fair merit pay system must find a way to accurately separate individual effort from team effort and reward each accordingly. As noted earlier, it is difficult to separate out such individual achievement from team achievement in joint production efforts. Finally, the authors note that the output of the educational system is difficult to measure, especially the output that the teacher is directly responsible for producing. Unlike the production of a good, one cannot merely count the output of the educational process. Although testing has improved over the past 25 years, it is still difficult to parse out exactly the impact that a teacher has on a student versus the impact of other influences. This is especially true if this ill-defined output is produced by a team that includes both employees and non-employees of the school system.

In addition, the principal-agent theory suggests that merit pay may be particularly difficult to implement in an educational setting (Goldhaber, DeArmond, Player, and Choi, 2008). According to this theory, a principal retains an agent to perform work that profits the principal. Unfortunately, there is typically imperfect information regarding the agent's efforts on behalf of the principal. Usually, the agent knows much more about the nature of his or her work than the principal does. This theory is very relevant to the field of education: school administrators (the principals) do not know how much a teacher works, the level of their effort, or the effectiveness of their practices; measures of output (graduation rates, test scores) are imperfect at best and are affected by many other factors besides the efforts of the teachers. Thus, it may be difficult to compensate teachers based on their efforts or output, especially if the cost of acquiring such information is very high. The next section discusses some of the recent research on the topic of merit pay and student academic attainment.

LITERATURE REVIEW

Most of the prior empirical studies done on the effectiveness of teacher attainment pay have used only regional data or data from foreign schools (Fryer, 2011; Glewwe, Ilias, and Kremer, 2010; Glazerman and Seifullah, 2010; Goodman and Turner, 2009; Lavy, 2009, 2002; Eberts, Hollenbeck, and Stone, 2002; Ladd, 1999). An excellent review of the empirical literature and a descriptive analysis of several merit pay plans being used in various schools and districts in the US are presented in Podgursky and Springer (2007).

Only one prior study looked at the effects of teacher merit pay using data from across the U.S. (Figlio and Kenny, 2007). In this study, the authors conducted a survey of both public and private high schools in the U.S. Their survey was mostly concerned with teacher merit pay at the

school district level. Only about 40 percent of schools responded to the survey; in addition, there was a disproportionate response rate from private high schools. This survey data on merit pay was then combined with data obtained from the National Education Longitudinal Survey (NELS) and the Schools and Staffing Survey (SASS). The author's survey was conducted in 2000, the NELS data was collected in 1992, and the SASS data was collected in 1993. The authors did not believe that this eight year gap was a cause for concern. However, because the merit pay data was from 2000 and the student attainment data was from 1992, there must be some type of incongruity between these two sets of data. Nonetheless, their results suggested that those schools that have merit pay for teachers have greater student achievement, where achievement is measured as 12^{th} grade test scores for math, reading, science, and history. However, even Figlio and Kenny (2007) admitted that they cannot determine if this relationship is causal or if better performing schools are just more likely to implement teacher merit pay programs.

Regarding other research in this area, Fryer (2011) examined the impact of teacher attainment pay on student attainment in New York City schools. A random sample of over 200 schools found that these incentives did not result in increased academic attainment as measured by grades, standardized test scores, attendance, and graduation rates. In fact, it was found that implementation of these merit-based pay standards actually reduced student achievement. Reasons given as to why merit pay didn't work include the following: incentives were too small; merit pay system was too complex; group-based awards were ineffective; and low response rate for merit pay paperwork. It was found that only 76 percent of teachers eligible for merit pay in 2009 completed the necessary paperwork to earn their bonus.

In Glewwe, Ilias, and Kremer (2010), a random sample of teachers in Kenya was used in order to determine if merit pay increases test scores and teacher attendance and reduces the student dropout rate. It was found that the incentive program increased students test scores but did not affect teacher attendance or the dropout rate. Unfortunately, the reason for these inconsistent results is because teachers focused solely on increasing students test scores, going so far as to hold test preparation sessions outside of normal school hours. In addition, given that test scores were weighted heavily in the teachers' merit pay system, teachers made little effort to lower the dropout rate or to increase their own attendance at school. Finally, there is no evidence that actual learning and greater acquisition of human capital occurred due to the existence of a merit pay system. Instead, it appears as if the teachers in Kenya ended up responding to the incentives and taught to the test.

Glazerman and Seifullah (2010) assessed the impact of the Chicago Public School's Teacher Advancement Program (TAP) on leaning outcomes. Under TAP, teachers could earn additional pay by being promoted to mentors or master teachers and could earn annual bonuses that are based on student attainment and observed classroom behavior. In its second year, TAP bonuses ranged from $2,600 to $6,320. The authors found no evidence that TAP increased students' test scores. In addition, teacher retention rates were not affected by the TAP program.

According to Glazerman and Seifullah (2010), these results are robust to the use of different samples and varying estimation techniques.

In Goodman and Turner (2009), a group-based merit pay system was examined for a subset of New York City schools. The purpose of this study was to determine if this type of merit pay increased student achievement. Results indicated that this group-based merit pay system had no statistically-significant effects on student achievement as measured by scores on annually administered math and reading exams.

Lavy (2009) looked at the effect of merit pay on English and math test scores in Israel. Student attainment was measured by scores on high school matriculation exams. It is important to note that the merit pay system lasted only one year and that the demographics of the student population in Israel are vastly different from that of the US. In addition, the sample selected was not random, and results from the difference-in-difference regression were biased. Nonetheless, however, Lavy found that merit pay did result in higher test scores in English and math. Lavy also found that the merit pay system enticed teachers to alter their teaching styles in order to increase test scores. Lavy, however, made no conclusions regarding the potential long term effects of this program, nor did he provide evidence that merit pay actually increased the acquisition of human capital beyond the increased test scores.

Lavy (2002) examined a merit pay system implemented in 62 schools in Israel in 1995. Results of his analysis suggested that merit pay had a significant and positive effect on student achievement, as measured by test scores and dropout rates. The results of this study were somewhat limited, however, due to the non-randomness of the sample of schools examined and the small samples sizes of some of categories of schools.

Eberts, Hollenbeck and Stone (2002) used data on two high schools in Michigan, one that utilized a merit pay system and another that used the traditional single salary schedule in order to determine if merit pay had any effect on student achievement. Employing a difference-in-difference estimation technique, the authors found that merit pay did not work; in the school that implemented the merit pay system, grade point averages fell, attendance fell, and course passing rates declined in comparison to the control school.

Finally, Ladd (1999) studied the effects of a attainment pay system on student student attainment in the Dallas, Texas school system. Using state-administered standardized tests on reading and math for seventh graders as a measure of academic attainment, Ladd found that the academic attainment of White and Hispanic students increased after the implementation of the merit pay system, while the attainment of African-American students was unaffected by the program. The reasons for this difference were unknown.

In summary, the results of prior research are mixed. Merit pay appears to work in Israel and Dallas, but not in New York, Chicago or Michigan. It worked to some extent in Kenya, but even there, certain outcomes were less than desirable. Most of the prior research looked at only small subsets of schools or schools in foreign countries. In several of these studies, the authors admitted that their samples are non-random. This study expands on this body of research by

using restricted-access data from the US Department of Education to determine if merit pay at the district level has any statistically-significant effect on student achievement. The next section presents the empirical model that will be used in this study.

EMPIRICAL TECHNIQUE AND DATA

In order to examine the relationship between merit pay and student attainment, it is assumed that academic achievement is dependent upon the demographics of the student body and the resources available to the educational institutions. By incorporating all of these factors, one takes account of all of the various agents responsible for teaching a child. Hence, districts that have above average teacher salaries, low student-teacher ratios, and a well-educated populace will, on average, have higher student achievement. Ladd (1999) utilized a similar estimating model in her examination of merit pay on student attainment in Dallas, Texas.

Regarding student attainment of academic goals, two measures will be used. The first measure of academic attainment is the average freshman graduation rate. This rate is the percentage of freshmen who graduate in four years. The second measure is the total dropout rate, which is the percentage of all students in grades 9 through 12 who dropout. It is important to note that the dropout rate is not one minus the graduation rate. Students may leave the school district for reasons other than dropping out. Hence, both measures are used in order to obtain a more accurate portrayal of student academic attainment at the district level.

The following equation will be estimated in the study:

$$Y = a_0 + a_1\,PAY + a_2\,BONUS + a_3\,STR + a_4\,WHITE + a_5\,BLACK + a_6\,HISPANIC + a_7\,COLLEGE + a_8\,HIGH + a_9\,URBAN + a_{10}\,RURAL + a_{11}\,DAYS + a_{12}\,NORTH + a_{13}\,SOUTH + a_{14}\,MIDWEST + a_{15}\,UNEMP + a_{16}\,GINI + a_{17}\,INCOME + a_{18}\,LUNCH. \quad (1)$$

The variables are defined as follows:

1. Y is a measure of student achievement. Two measures are used in the study: (a) average freshman graduation rate; (b) average drop-out rate.
2. PAY is the average district-level salary for public school teachers
3. BONUS is a dummy variables that equals one if the district rewards excellence in teaching and zero otherwise
4. STR is the district-level student-teacher ratio
5. WHITE is the percentage of the student population that is White
6. BLACK is the percentage of the student population that is African-American
7. HISPANIC is the percentage of the student population that is Hispanic
8. COLLEGE is the percentage of the district's population whose highest educational attainment is a bachelor's degree

9. HIGH is the percentage of the district's population whose highest educational attainment is a high-school diploma
10. URBAN equals one of district is located in an urban area and zero otherwise
11. RURAL equals one if district is located in a rural area and zero otherwise; the excluded category for URBAN and RURAL is suburban, as defined in SASS.
12. DAYS is the length of the school year in days
13. NORTH equals one if the district is located in a Northeastern state and zero otherwise
14. SOUTH equals one if the district is located in a Southern state and zero otherwise
15. MIDWEST equals one if the district is located in a Midwestern state and zero otherwise
16. UNEMP is the district's unemployment rate
17. GINI is the district's Gini coefficient. The Gini coefficient is a measure of the income inequality in the district. Zero indicates total income equality; one indicates maximum inequality.
18. INCOME is the median family income
19. LUNCH is the percentage of the district's student who obtained reduced fee or free lunches.

Regarding the BONUS variable, no information is available on the amount of financial incentive involved, nor on how "excellence in teaching" is defined by each individual school district. All that is known is whether or not the district in question rewards excellence in teaching in some monetary fashion. It may very well be that one district's definition of excellence is not the same as another district's definition of excellence. Also, it is not known if merit pay takes the form of a one-time bonus or a permanent increase in the base pay of the teacher. Even though this binary variable does not shed any light on the type of financial bonus used to reward excellence in teaching, it should nonetheless be a reasonable proxy for the existence of a teacher merit pay system in a given district.

Median family income, the Gini coefficient, and percentage of students receiving free or reduced-fee lunches are included in order to capture any possible effects of family income on a student's capacity to learn. It is reasonable to assume that students whose families earn higher incomes are more likely to have parents who are college-educated, are more likely to have parents who value education, and are more likely to have opportunities, like travelling to foreign lands, that enhance the learning received in school. Hence, the greater the median family income of a district, the higher its graduation rate should be and the lower its dropout rate should be, holding all else constant.

The Gini coefficient, an index of income inequality, is included in order to capture possible public school flight. If a district has a very large amount of income inequality, it may cause upper-income families to send their children to private schools. Thus, a district may have a relatively high median family income, but if there is a large amount of income inequality, public schools may be populated with predominately lower-income children. In order to test for that possibility, the percentage of students receiving free or reduced-fee lunches is also included as an explanatory variable. This LUNCH variable is a proxy for the median-family income of students enrolled in district schools. Hence, if median district-level income is high, but if there is substantial income inequality and many students obtain free or reduced-fee lunches, then many

upper-income families may be sending their children to private schools. This may then result in lower levels of student attainment in district schools even though district-level median family income is relatively high. It is important to note, however, that this particular type of district (high median family income, high percentage of students qualifying for free or reduced fee lunches, and high Gini coefficient) is very rare; in looking at those districts with median family income above $100,000, less than 10 percent have free lunch percentages above 43 percent, the national average.

Racial variables are included in order to determine if the racial composition of a district has any impact on graduation or dropout rates. It is expected that those districts with larger percentages of minority populations will have lower graduation rates and higher dropout rates (Ladd, 1999). Regional and urban variables are included in order to capture any possible academic attainment differences between different regions of the country and between urban and rural areas. Educational attainment variables are included in order to capture the proclivity of the district's populace to fund educational initiatives (Figlio and Kenny, 2007).

All of the education-related data were obtained from either the restricted access version of the Schools and Staffing Survey (SASS) of the U.S. Department of Education or from various publicly-available surveys that are contained in the National Center for Education Statistics; district-level demographic and socioeconomic data were obtained from the American Community Surveys. All data is district level for the academic year 2007-2008. All educational data is for public schools only. Data on approximately 1,620 (rounded to nearest 10) school districts were included in the final sample; sample size was approximately due to use of restricted data.

One potential problem with the study is its use of district level data. A disadvantage of using district-level data is that some data that is available at the state level or school level is not available at the district level. For example, data on assessment tests, such as the National Assessment of Educational Progress (NAEP), are only available for a very limited number of districts. Data on NAEP scores are readily available, however, at the state level for all states. In addition, the use of school-level data would allow for the use of much richer instruments of student achievement and teacher efforts, such as grade point averages and hours worked in the classroom. However, an advantage of using district level data is that much more readily available demographic and socioeconomic data are available at the district level than at the school level. In addition, the use of district level data avoids some of the aggregation problems associated with state-level data.

RESULTS

Descriptive statistics are presented on Table 1.

Table 1: Descriptive Statistics N=1620 (rounded to nearest 10)		
Variable	Mean	Standard Deviation
Graduation Rate	76.26%	13.25
Dropout Rate	3.73%	2.94
PAY	$51,950	8295
BONUS (Excellence in Teaching)	10.12%	30.18
BONUS (Certification)	52.4%	49.9
STR	14.74	4.11
WHITE	64.25%	27.10
BLACK	16.2%	20.5
HISPANIC	14.74%	20.03
COLLEGE	27.65%	14.1
HIGH	59.7%	8.73
URBAN	23.3%	42.3
RURAL	19%	39.2
DAYS	178	3.3
NORTH	16.55%	37.17
SOUTH	40.27%	49.06
MIDWEST	23.5%	42.47
UNEMP	4.9%	2.08
GINI	0.427	0.044
INCOME	$64,857	21117
LUNCH	40.98%	21.85

According to these statistics, 10.1 percent of districts had a merit pay plan for teachers that rewarded excellence in teaching, the average graduate rate was 76.26 percent, the average dropout rate was 3.72 percent, the average student teacher ratio was 14.7, and 40 percent of the students received free or reduced price lunches. In comparing that percentage to the general population of students, 43 percent of all students in the U.S. received a free or reduced price lunch; hence the sample used in this study is very representative of the population, at least with regards to the percentage of students receiving free or reduced fee lunches. Finally, it is important to note that a sizeable minority of the districts are from the South (40 percent). Although graduation rates have typically been lower and dropout rates have typically been higher in the South, it is unclear if this over-sampling from the South would introduce any bias into the regression analysis. By including regional dummy variables in the regression, regional variations should be taken account of and any potential bias that this over-sampling may have introduced would have been mitigated due to the use of these regional variables.

In addition to the multivariate analysis, a simple t-test was used to compare the means of the graduation and dropout rates for those schools who had merit pay and for those schools who did not have merit pay. The test statistic for the graduation rate test was -4.41, while the test statistic for the dropout rate test was 1.48. The first test statistic suggests that the average graduation rate was statistically lower in districts that had merit pay than in districts that did not have merit pay. The latter test statistic indicates that there is no statistically-significant difference between dropout rates in districts that had merit pay and in districts that did not have merit pay. These results may suggest that either merit pay doesn't work or that high performing districts did not feel a need to implement it.

In order to test this theory, data was obtained on merit pay plans for the year 2004. There were approximately 80 districts that did not have a merit pay system in 2004 but had implemented one by 2007. It was found that there was no statistically significant difference between the average graduation and dropout rates in 2004 and the average graduation and dropout rates in 2007 for this group of districts. The test statistic for the graduation rate was 1.044, and the test statistic for the dropout rate was -0.975. Although not conclusive evidence, these results suggest that merit pay may not be effective in increasing district-level student achievement.

Ordinary least squares (OLS) was used to estimate equation (1). Regression results are presented on Tables 2 and 3. These results suggest that merit pay is not significantly related to either of the measures of student attainment that were examined. This finding is significant since the results of recent research on this topic have been mixed. Some prior studies have found that merit pay plans had significant effects on student attainment (Lavy, 2009, 2002; Glewwe, Ilias, and Kremer, 2010; Figlio and Kenny, 2007; Ladd, 1999). Many other studies, however, found that merit or attainment pay plans had no effects on student achievement (Fryer, 2011; Glazerman and Seifullah, 2010; Goodman and Turner, 2009; Ebberts, Hollenbeck, and Stone, 2002).

Regarding the significance of other variables, districts that had larger percentages of African-American students had lower graduation rates; for every one percentage point that the share of African-American students increased by, the graduation rate dropped by 0.26 points. Districts with larger percentages of African-American students had higher dropout rates. Districts with larger percentages of residents with a college degree had higher graduation rates and lower dropout rates; for every one percentage point that the share of college-educated adults increased by, the graduation rate increased by 0.54 points, and the dropout rate fell by 0.077 points. Districts with higher unemployment rates had lower graduation rates; for every point increase, the graduation rate fell by 0.23. Higher percentages of students receiving free or reduced fee lunches resulted in lower graduation rates; for every point increase, the graduation rate fell by 0.041 points. Districts with lower incomes had lower graduation rates and higher dropout rates; for every $1000 decrease in median family income, the graduation rate fell by 0.092 points, and the dropout rate increased by 0.022 points. These results are reasonable and

suggest that many factors not under the control of the teachers or school administrators have rather large effects on student achievement (Figlio and Kenny, 2007; Ladd, 1999).

Table 2: OLS Regression Results
Dependent Variable: Graduation Rate

Variable	Coefficient	Standard Error	Test Statistic
Constant	26.49	14.45	1.833*
PAY	0.000027	0.000036	0.767
BONUS (Excellence in Teaching)	-0.829	0.707	-1.173
STR	0.039	0.0578	0.677
WHITE	-0.0318	0.0331	-0.962
BLACK	-0.2597	0.0346	-7.511***
HISPANIC	-0.04711	0.0351	-1.344
COLLEGE	0.5418	0.0524	10.327***
HIGH	0.3568	0.0549	6.49***
URBAN	-1.582	0.593	-2.668***
RURAL	2.855	0.639	4.462***
DAYS	0.0943	0.069	1.367
NORTH	4.95	0.89	5.544***
SOUTH	5.77	0.761	7.575***
MIDWEST	7.27	0.754	9.637***
UNEMP	-0.2344	0.1371	-1.710*
GINI	-15.33	6.922	-2.215**
INCOME	0.000092	0.000025	3.665***
LUNCH	-0.0411	0.01745	-2.355**

Notes:
Adjusted $R^2 = 0.604$
F = 138.43
10% Level of Significance = *
5% Level of Significance = **
1% Level of Significance = ***

In order to test the robustness of the above results, another type of merit pay was examined. One of the variables included in the SASS data set was a dummy variable that equaled one if merit pay was awarded to a teacher who became certified through the National Board for Professional Teaching Standards (NBPTS) and zero otherwise. This certification is an advanced teaching credential that teachers obtain after undergoing a thorough assessment of their teaching abilities. In the sample used in this study, 52 percent of districts gave some type of merit pay to teachers for obtaining NBPTS certification. Once again, it is important to note that

it is not known in what form the merit pay was awarded. However, unlike the "excellence in teaching" criteria for awarding merit pay, there is much less subjectivity with this criteria. In all districts that awarded this type of merit pay, if a teacher had earned NBPTS certification, then they would be eligible for the merit pay. In order to see if this type of merit pay resulted in an increase in student academic attainment, equation (1) was re-estimated, replacing the "excellence in teaching" merit pay dummy variable with an NBPTS merit pay variable. The NBPTS variable equaled one if the district awarded merit pay to teachers who had earned the NBPTS certification and zero otherwise. Results are presented on Tables 4 and 5.

Table 3: OLS Regression Results
Dependent Variable: Dropout Rate

Variable	Coefficient	Standard Error	Test Statistic
Constant	25.744	4.32	5.946***
PAY	-0.0000046	0.000011	-0.431
BONUS (Excellence in Teaching)	0.203	0.212	0.962
STR	-0.0361	0.0173	-2.083**
WHITE	-0.01268	0.0099	-1.28
BLACK	0.02162	0.01035	2.087**
HISPANIC	-0.0122	0.0105	-1.159
COLLEGE	-0.07754	0.0157	-4.933***
HIGH	-0.05313	0.01647	-3.226***
URBAN	0.456	0.177	2.569**
RURAL	-0.383	0.192	-1.998**
DAYS	-0.0822	0.0206	-3.979***
NORTH	-1.039	0.267	-3.884***
SOUTH	-1.803	0.228	-7.907***
MIDWEST	-1.62	0.226	-7.209***
UNEMP	0.0323	0.04106	0.787
GINI	4.358	2.074	2.101**
INCOME	-0.000022	0.0000076	-2.938***
LUNCH	0.00173	0.00522	0.331

Notes:
Adjusted $R^2 = 0.282$
F = 36.41
10% Level of Significance = *
5% Level of Significance = **
1% Level of Significance = ***

Results suggest that merit pay based on national teacher certification is significantly and negatively related to graduation rates and positively related to dropout rates. These results suggest that certification in and of itself may not increase student academic attainment. Another possible explanation for this finding is that high performing districts may not feel the need to reward teachers that obtain NBPTS certification. Hence, only low performing districts may offer this type incentive, thus creating the perverse result we see here.

Table 4: OLS Regression Results
Dependent Variable: Graduation Rate

Variable	Coefficient	Standard Error	Test Statistic
Constant	25.82	14.36	1.798*
PAY	0.000031	0.000036	0.878
BONUS (Certification)	-1.96	0.447	-4.376***
STR	0.018	0.0575	0.314
WHITE	-0.0323	0.0328	-0.985
BLACK	-0.2544	0.0344	-7.397***
HISPANIC	-0.0546	0.0348	-1.568
COLLEGE	0.5428	0.0521	10.408***
HIGH	0.3513	0.0547	6.425***
URBAN	-1.587	0.589	-2.693***
RURAL	2.78	0.636	4.369***
DAYS	0.109	0.0687	1.59
NORTH	4.07	0.912	4.461***
SOUTH	5.552	0.752	7.38***
MIDWEST	6.665	0.762	8.746***
UNEMP	-0.2886	0.1368	-2.11**
GINI	-14.25	6.88	-2.068**
INCOME	0.000088	0.000025	3.498***
LUNCH	-0.0433	0.01737	-2.494**

Notes:
Adjusted $R^2 = 0.609$
F = 140.95
10% Level of Significance = *
5% Level of Significance = **
1% Level of Significance = ***

In fact, if we look at a simple t-test comparing the average graduation rate of those districts that reward certification and those that do not, we obtain a test statistic of -8.865. This result indicates that districts that reward certification have statistically lower graduation rates

than those districts that do not reward certification. For the dropout rate, the test statistic is 3.89, which indicates that districts that reward certification have higher dropout rates than those districts that do not reward certification. These results corroborate the results of the regression analysis.

Most of the other explanatory variables in the certification regressions were significant with the same signs as were found in the "excellence in teaching" regressions. These additional regressions illustrate the robustness of the results of this study in that merit pay, regardless of the criteria used to award it, does not improve academic attainment.

Table 5: OLS Regression Results
Dependent Variable: Dropout Rate

Variable	Coefficient	Standard Error	Test Statistic
Constant	25.76	4.32	5.963***
PAY	-0.0000054	0.000011	-0.499
BONUS (Certification)	0.289	0.135	2.148**
STR	-0.0325	0.0173	-1.88*
WHITE	-0.0124	0.00989	-1.261
BLACK	0.0209	0.01035	2.029**
HISPANIC	-0.01078	0.01047	-1.03
COLLEGE	-0.0775	0.0157	-4.938***
HIGH	-0.0522	0.01645	-3.174***
URBAN	0.459	0.177	2.594***
RURAL	-0.373	0.192	-1.948*
DAYS	-0.0842	0.0206	-4.074***
NORTH	-0.91	0.274	-3.317***
SOUTH	-1.762	0.226	-7.781***
MIDWEST	-1.538	0.229	-6.708***
UNEMP	0.04056	0.0411	0.985
GINI	4.22	2.073	2.036**
INCOME	-0.000021	0.0000076	-2.841***
LUNCH	0.00202	0.00522	0.387

Notes:
Adjusted $R^2 = 0.284$
F = 36.70
10% Level of Significance = *
5% Level of Significance = **
1% Level of Significance = ***

CONCLUSION

Using district-level data, the results of this study suggest that there is no statistically-significant and positive relationship between teacher merit pay and student academic attainment as measured by graduation and dropout rates; for certification bonuses, results suggest that such merit pay may even be related to lower student achievement. As noted earlier, one possible explanation for these results may be the principal-agent theory (Goldhaber, DeArmond, Player, and Choi, 2008). Given that there is imperfect information regarding the agent's efforts on behalf of the principal, it may be difficult to compensate teachers based on their efforts or output, especially in the education field due to the high cost of acquiring such information. Thus, given the difficulty in evaluating the direct impact that individual teachers have on student academic attainment and the logistical problems involved with rewarding effective practices and behavior appropriately, it is not surprising to find that there is no positive and significant relationship between merit pay and student academic attainment. Finally, this study is an important contribution to the body of research in this area because it is the first study to use a national set of district-level data, and its findings are consistent with some of the research in this area in that it finds that merit pay has no positive and significant effects on student academic attainment.

REFERENCES

Ballou, D (2001). Pay for Attainment in Public and Private Schools. *Economics of Education Review*, 20(1), 51-61.

Ebberts, R., K. Hollenbeck, and J. Stone (2002). Teacher Attainment Incentives and Student Outcomes. *The Journal of Human Resources*, 37 (4), 913-927.

Figlio, D. and L. Kenny (2007). Individual Teacher Incentives and Student Attainment. *Journal of Public Economics*, 91, 901-914.

Fryer, R. (2011). Teacher Incentives and Student Achievement: Evidence From New York City Public Schools. NBER Working Paper Series, Number 16850.

Glazerman, S. and A. Seifullah (2010). An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report. Mathematica Policy Research Reference Number: 6319-520.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, 2: 205-227.

Goldhaber, D., M. DeArmond, D. Player, and H. Choi (2008). Why Do So Few Public School Districts Use Merit Pay. *Journal of Education Finance*, 33 (3): 262-289.

Goodman, S. and L. Turner (2009). Group Incentives for Teachers: The Impact of the NYC School-Wide Bonus Program on Educational Outcomes. Department of Economics Discussion Paper no. 0910-05, Columbia University.

Ladd, H. (1999). The Dallas School Accountability and Incentive Program: An Evaluation of Its Impact on Student Outcomes. *Economics of Education Review*, 18, 1-16.

Lavy, V. (2002). Evaulating the Effect of Teachers' Group Attainment Incentives on Pupil Achievement. *The Journal of Political Economy*, 110 (6), 1286-1317.

Lavy, V. (2009). Attainment Pay and Teachers' Effort, Productivity, and Grading Ethics. *American Economic Review*, 99 (5): 1979-2011.

Murname, R. and D. Cohen (1986). Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and Few Survive. *Harvard Education Review*, 56 (1): 1-17.

Podgursky, M. and M. Springer (2007). Teacher Attainment Pay: A Review. *Journal of Policy Analysis and Management*, 26 (4): 909-949.

CONTRIBUTIONS FOR AN ECONOMIC PERSPECTIVE ON SCHOOL CHOICE

Ricardo Manuel Santos, Trinity University[1]
James Nordlund, Trinity University

ABSTRACT

This paper provides an introduction to school choice analysis from the perspective of economics. The contributions from different fields and researchers are compiled and organized in a thematic manner. We first define the issue and highlight the advantages of an economic perspective. The brief history of the school choice debate is introduced and the most influential contemporary authors presented. We discuss several theoretical arguments both in favor and against school choice from the schools and districts perspective (supply side), and students-parents view (demand side). The most important results from existing experiments are also summarized. We conclude there are two fundamental issues the analysis should incorporate: how information about school choice is disclosed to parents and schools, and how a proper evaluation about the programs is necessary. About the latter, we emphasize the necessity to select a proper control group and allowing enough time go by before analyzing the effectiveness of the programs being evaluated.

JEL codes: I22, I28, A11, A12, H42

BACKGROUND

Defining School Choice

School choice refers to the freedom of parents and students to select the students' form of education. While few would argue that our children ought not to be taught at all, the question of *how* we education our children is a much more pressing issue. Proponents of school choice tend to agree that parents and students should have more answers available to answer this question. In contrast, those arguing against this system find that making more schooling options accessible creates some larger net cost. While there are numerous models across the international educational system that claim to offer some form of "additional school choice" over the status quo, how these plans contribute to choice, if they do at all, vary greatly. It is therefore prerequisite to understand the jargon of the school choice debate before reviewing various policy models. It is our goal to provide an overview of the contributions for an economic perspective on school choice.

Currently, the typical methods by which parents select schools are choosing where to live (this determines the public school district to which they are a part of) or to attend a private alternative. Students receive education in one of four ways. In scenario (1), students are presented with privately financed and privately provided education (like private schooling or home schooling). Alternatively, in a different scenario, (2), students benefit from public financing but receive privately provided education (as is the case in school vouchers or charter schooling). In a third scenario, (3), students receive private financing to attend a publically provided school (this is more likely the case in tertiary education than in primary or secondary education). Finally, in (4), students attend a publically financed and provided school, such as the traditional public school. When compared to the situation in (4), more school choice implies a weaker governmental role. In (4), the government finances students' education at public schools. The idea of offering more freedom in how parents select schooling for their children does not mean we should reduce the public endowment in any way (so long as we are careful to prevent financing from influencing provision) and thus a shift towards case (1) or (3) alone would not make sense. School choice can be offered by increasing private provision – a shift from (4) to (2) – or by a restructuring of the status quo of the public educational system – (4) to (4) – by creating some change that reduces current student and school restrictions. Regardless of whether we move from (4) to (2) or from (4) to (4) with restructuring, we maintain the idea of public provision, and thus are required to discuss the justifications for governmental intervention.

Justifications for governmental intervention in a market tend to come from two concepts: efficiency and equity. The efficiency argument relies on the existence of market failures; hence, the government is the appropriate agent to step in and improve efficiency. Two important potential market failures in education are the existence of externalities or incomplete information. Because a student's education increases their human capital (making them a more productive worker), education is commonly viewed as a positive externality. Given the education of a student creates an unrecognized or under-recognized benefit to society the government could justify intervention in the education market. A second important market failure is the problem of incomplete information. Parents and students are unable to measure properly the long-run value of education, which could lead to a less than optimal demand for education. The government's role is thus to provide the necessary level of education. The government can also justify intervention in the education market because of equity. This argument is rooted in the idea that inequity of family resources between families should be corrected (e.g., families with greater resources have a higher ability to finance a private alternative or to choose to live in a better school district). Hence, this justification is more closely related to issues of public financing than of public provision. Either an inefficient or an inequitable market could justify the existence of public financing, though the debate is more intense over the issue of public provision. This explains why reactions are generally milder when considering a restructuring in scenario (4) to (4) than in moving from (4) to (2).

Why an Economic Perspective

Economists are well equipped to solve a variety of problems using tools developed for use in economic analysis. For this reason, economists have a lot to offer the policy debate regarding school choice. A principal tool used in the literature surrounding school choice is an analysis of market structure. One cannot advocate a change in school choice opportunities without considering the effect such a change will have on the market structure for education. It is necessary to study this new system since changes in the market structure will influence parents and students' behaviors. This in turn will affect outcomes such as school achievement, productivity, and parental satisfaction. Economists are well practiced in studying how agents respond to market incentives so they can identify and measure the relationship between changes in the market structure and outcomes. This discussion of market structure considers not only the final good (education, as achieved through an interaction between students and schools) but also the labor market for teachers and school administrators. Teachers are the most important input in the production of education, and the market for teachers is strongly affected by the implementation of school choice programs, especially given concerns regarding the level of salaries at different schools. More generally, economists are well equipped to deal with a wide variety of supply side factors such as input costs, economies of scale, and school attributes which parents pay for. A study of how these inputs generate education is nothing more than a description of the production function for schooling. Since much of the funding for primary and secondary education comes from the public coffer, researchers are also frequently asked to discuss the efficiency of inputs used in the described production function. An economic approach to school choice also allows researchers a methodology for identifying the total effects of a policy change on society. Ideas like student segregation ("cream-skimming") require a generalized sample analysis. Focus on individual behavior is not sufficient given the asymmetry of the results may be large and variable with the socio-economic characteristics of the students. A general equilibrium model allows for an analytical structure that relates all students from all schools, and finds the total surplus of school choice for society.

Two Fundamental Notes About School Choice

Poorly informed parents and students will make decisions that differ from what a fully rational individual would make. This leads to an increase in the inefficiency of the educational system. If the market for education works well, then schools and school districts will have the appropriate incentives to collect and distribute *information* to its consumers. However, given the difficulty in quantifying the quality of schools, mentioned above, schools may attempt to falsify quality reports to make their school appear better than it is. Existing quality regulations prove schools indeed fall to this temptation.[2] Information also has implications with respect to inequality: lower income families have greater difficulty obtaining information regarding the

quality of schools. Consequently, they are more likely to make bad choices. Thus, a primary concern with information is making it available and accessible to everyone, with particular focus on those that would otherwise not have access to it.

There is also a fear that, given the quality of schools is difficult to measure, analysts will work with inaccurate information and thus advocate a school choice reform model that is unfounded. There is no simple *evaluation* of school choice experiments. Merrifield (2008) argues that existing small-scale experiments in the United States rarely do more than tweak the status quo – thus indicating results will have low significance – and ignore many of the potential issues that should arise under a system with greater school choice. He writes, "Several key aspects of market accountability are virtually absent from those programs: price change, easy market entry, and the profit motive, among others." Some of these topics are likely less frequent in the literature about school choice because these evaluations are more difficult to make. In section 4 we describe the underlying parameters that should be considered in an ideal evaluation scenario.

The next section describes the origin of school choice. Section 3 summarizes the arguments in favor and against school choice from both perspectives of students (as well as the parents) and schools (as well as the districts). In the following section, we describe the ideal evaluation scenario for a school choice experiment as well as the results from existing experiments. Section 5 concludes.

ORIGIN OF SCHOOL CHOICE

Many economists and non-economists have contributed to the debate surrounding the expansion of freedom of choice. Friedman (1955) was the first to propose reforms to increase school choice. He recognized that schooling created public benefits that outweighed individual costs and thus justified governmental intervention. An important contribution by Friedman was to recognize the necessity of differentiating public provision and public financing. Friedman suggested that school vouchers would separate public provision from public financing, giving parents more freedom over which schools their children attended, thus stimulating competition in the supply of education. Two other early researchers of school choice and educational vouchers are Jenks and Fantini, though the two have differing opinions as to how education should be provided. Jenks (1966) suggested that private schools would be a strong alternative to existing public schools. He stressed one of the principal failures of the public school district came from its bureaucratic system. Thus, he suggested the use of vouchers and private provision. He designed the Alum Rock experiment in California. The school was overregulated and so it achieved weak results. Fantini (1973) favored the public provision of education. His idea was to reform the public system to better accommodate students with different learning styles and interests. Initially his ideas were very influential. A more discriminating approach to use of vouchers was first proposed by Coons and Sugarman (1978). The two focused their attention on

equity and argued that education vouchers should be of different amounts, depending on tuition, family income, and the willingness of the family to invest in education. The main difficulty with this type of voucher is that it required a large amount of information.

Not all researchers have been narrowly focused on issues of financing. For example, Chubb and Moe (1990) provide a good overview of the effect of school choice on bureaucracy in public schools. They believe that school's ability to independent decision making is the most important reason for student achievement in schools. They propose a new system in which centralized bureaucracy is eliminated and more authority is given directly to schools, parents, and students. As more schools across the United States implemented school choice programs, researchers were able to perform a more technical analysis of school choice. Hoxby (2000), Greene and Forster (2002), West and Peterson (2006), Epple and Romano (2002), and Hanushek, Klain, Rivkin (2001) are some of the most influential contemporary authors. They provide valuable examinations of existing programs (their flaws and successes) through rigorous statistical analysis.

ECONOMIC PERSPECTIVE ON SCHOOL CHOICE – THEORY

There are numerous arguments both in favor of and against school choice. These arguments are categorized as supply side (schools and districts) or demand side (students and parents) arguments and are summarized below.

Favor – Supply Side

One argument in favor of expanding school choice is that the existing system is characterized by heavy regulations, which leads teachers and administrators to spend valuable time ensuring requirements are met. Peterson (1990) argues that teachers and administrators should be less concerned with paperwork and more focused on finding new teaching methods that would increase the quality of education and better meet students' needs. Additionally, since a less bureaucratic system allows for an increased independence of schools regarding curriculum, teacher contracts, and teaching techniques, there will also more flexibility in finding solutions that meet students' demands. McCluskey (2005) finds that the threat of school choice (through increased competition) is more efficient and effective than bureaucracy in regulating a school's behavior. Forster (2009) provides an interesting synthesis of studies on the impact of vouchers on public schools and finds that vouchers improved the quality of public schools in 16 of the 17 studies considered.

Others argue that increasing school choice will lead to a more efficient production of education (providing the same educational quality at a lower cost). Salisbury (2005) finds that existing programs dedicated to expanding school choice in Arizona, Milwaukee, Cleveland, Florida, Pennsylvania, Maine, and Vermont do not impose an extra financial burden, and

analyses of proposed school choice programs in Utah, South Carolina, New Hampshire, Baltimore, and Virginia indicate that those programs would save the local or state government money. Coulson (2008) finds that tuition tax credits (a method to expand school choice) in New York, Illinois, and Texas eventually save these states as much money in the long run as they cost to implement. Tuition tax credits have also been found to increase education quality; thus, such a program could lead to a more efficient production of education (Forster and D'Andrea, 2009).

Favor – Demand Side

Supporters of school choice also consider a range of potential benefits that such a market would provide students and teachers, one of which is better matching between students and schools. After an initial adjustment period, parents will identify a school capable of satisfying their students' needs. Advocates of school choice assume this will occur because parental participation will increase as freedom of choice expands. For example, under some school choice proposals, a student's residence would no longer restrict the number of schools in the city to which he could attend. Thus low-income parents with more limited housing options now have other alternatives to decide where their children go to school. Forster (2005) gives an important condition for parental participation: he shows that when providing opportunities to participate in a school choice program - such as a voucher - are not extremely difficult, parents are more likely to be involved in exploring school choice options.[3] This hypothesis is corroborated by the fact that parents do indeed make decisions on schooling that are sensitive to school performance (Greene, 2010).

Other arguments are focused on the advantages of biased financing. Programs in which the amount of funding a student receives is inversely related to his/her family's income allow for a more equitable opportunity for education. Peske and Haycock (2006) claim that a significant teacher quality inequality exists between higher-income and lower-income schools. Deming (2009) defends that the benefits of a better education extend beyond the classroom; for example, lottery winners who were allowed to attend better schools in the Charlotte-Mecklenburg school district committed fewer and less serious crimes than other low-income students who remained in the lower-quality school. Further, biased financing provides schools with a diverse student body. Exposure to students from different backgrounds has positive effects on students, particularly those in urban areas. Finnigan and Stewart (2009), supporters of interdistrict school choice, find that such a program increases minority enrollment in suburban schools; as a result, these students display greater professional aspirations and lower social isolation than students that remain in urban districts.

Against – Supply Side

Not every researcher agrees that school choice will improve the market for education, and some that criticize it do so because it will be a detriment to the supply of education. First, challengers of the school choice movement fear that such a system facilitates discrimination by schools. This comes in either the form of cream skimming or overspecialization. In the first case, cream skimming may occur if schools are given a say in the selection of students. For example, if schools are allowed to select students based on ability, competition leads to stratification by parental income and reduced student effort (MacLeod and Urquiola, 2009). Romano and Epple (2002) design alternative voucher schemes and conclude that in order to avoid cream skimming the voucher system should have two characteristics: the voucher should be variable with a student's ability and schools must accept the voucher as the only way to pay tuition (no "top-ups" from higher income families to ensure their students a place). In the latter case, since schools will exist in a free market – provided the quantity of schools is sufficiently large – the need to attract students may encourage unfocused curriculums that are biased towards a specific type of student, creating overspecialized schools (Fiske and Ladd, 2000).

Second, even with no discrimination by schools, it may still be difficult for students to find the best program to meet their needs. Students may not identify the best school for the form of education they desire immediately, or overzealous parents with low performing students may use the school as an excuse for lazy work, thus transferring these students from school to school indefinitely. Hanushek et al. (2001) find that students in schools with higher turnover are disadvantaged compared to students in schools with lower turnover.

Finally, implementing school choice programs is no small feat, even assuming that students can quickly identify the best school to fit their needs and that no discrimination occurs. Critics of school choice worry that the cost of such a reform will be extraordinarily high. While some small experiments have found school choice programs to be fiscally neutral at worse for state or local governments, these may not consider every cost associated with school reform. Wells (1993) shows that in addition to funding vouchers or tax-credits, other more difficult to measure costs exist. One such cost is that as parents move their children away from the neighborhood public school, transportation costs increase (students must travel further to get to school, implementation of a city-wide school bus route to cater to every student's needs is infeasible). Moreover, is expected that increased competition for students forces schools to collect more (costly) information that identifies and sells to potential buyers.

Against – Demand Side

Others that criticize school choice worry that it will negatively affect students in some way. Greater school choice could marginalize students with special needs. This argument is founded on the belief that schools will not be willing to invest in infrastructure that applies only

to a minority group of students. Existing special needs programs in school choice models – for example, the Georgia Special Needs Scholarship or Florida's McKay Scholarship Program – indicate that this concern may be unrealistic.[4,5] Other critics of school choice worry that parents and students will make decisions based on extracurricular factors. Cullen et al. (2006) discuss a parent choosing to send their child to a different school for safety reasons, or a student choosing to attend a different school for extracurricular activities. They acknowledge, "Although these reasons may lead to improved life outcomes in the long run, they are less likely to influence traditional academic achievement measures in the short run.". Nevertheless 85% of parents state they consider academic quality a very important factor in deciding where to send their children, this concern may be unfounded (Peterson, 1999).[6] Regardless, it does not seem unreasonable to think that some students may be tempted by extracurricular factors (for example, a star athlete electing to attend a school with a strong athletics program).

All these arguments in favor and against present valid concerns. The final outcome of a School Choice program will be a positive one if the benefits outweigh costs. To properly evaluate the programs it is important to set the background parameters one should not ignore. We turn to that discussion below.

ECONOMIC PERSPECTIVE ON SCHOOL CHOICE – PRACTICE

The Ideal Evaluation

There are certain prerequisites to establishing conclusive findings in school choice research. The choice of proper control groups and adequate sample time are two imperatives for an economic analysis of school choice. The existence of an appropriate *control group* is necessary to identify any causality between availability of school choice and increased school quality (Campbell and Stanley, 1963). Every student and teacher has his own unique characteristics; the same is true for every school, district, city, and state. Researchers can only perform statistical inference when granted a proper control group to which they can compare students who have access to school choice to those who do not, holding all other variables relatively constant. From a public policy perspective, this is extremely difficult to achieve. Because different districts and cities possess different traits, it is not appropriate for researchers to compare a city with a municipal-wide school choice program to a city without such a program. However, granting access to only a portion of a local student body, while holding other students in the status quo (a result achieved by lotteries) often produces cries of injustice by parents of students left unlucky in the lottery. Also worth noting is that in a system that utilizes random lotteries, it is also important to consider whether such a lottery produces a significant changes for families' budget constraints. A student from a high-income family could apply for a tuition voucher and have his application rejected, but that does not imply his access to a quality

education is limited in any way. Thus, the ideal evaluation must consider policy that is directed specifically at students that have the most potential to gain from such a reform.

Second, the ideal evaluation of a school choice program requires that a school choice program remain *active for years before results are considered*. Part of this is due to the fact that economists require a plethora of observations, given the array of cross-sectional variables affecting results. Moreover, parents and students may be slow to react to a policy change – or at least to reach market equilibrium. It will also take time for schools to identify their optimal strategies for recruiting students, and it will take time for parents and students to learn how to use the school choice system to maximize their education per their needs.

Important Results

First and foremost, the objective of increasing school choice is to improve student achievement. Various studies of school choice programs find a positive relation between increased freedom of choice and student achievement.[7] Greene et al. (1998) show that test results in Milwaukee indicate that after four years, students who were awarded vouchers through a lottery had reading scores 6 NCE points higher than students who applied for vouchers but did not receive them; voucher students also had math scores 11 NCE points higher.[8] Cowen (2007) considers a similar program in Charlotte and finds that after only one year of voucher program implementation, voucher students had reading scores 8 percentile points higher than the control group, and math scores 7 percentile points higher. Barnard et al. (2003) study the New York randomized voucher program and show voucher students had math scores 5 percentile points higher than the control group. Wolf et al. (2007) review the voucher program in Washington D.C. and suspect that achievement gains appear likely for voucher students, though the results were statistically significant with only 93% certainty. Howell and Peterson (2002) provide a more narrow study focused on black voucher students and were able to find with statistical significance that these voucher students had combined reading and math scores 9 percentile points higher than the control group. Thus in general, it appears that school choice programs have been able to increase student achievement.

Second, economic analysis can and should evaluate the impact of the government's role in education. This portion of the analysis ought to consider whether public financing under the reformed system is efficient and whether a more competitive market with funding programs for lower income families generates greater equity in education. Results indicate that school choice encourages academic improvement and increased school quality. The main concern is whether public financing of school choice programs impose additional costs on state governments. This does not appear to be the case. Aud (2007) reviews existing voucher and tax-credit scholarship programs and finds that school choice programs have saved the states $22 million and districts $421 million. She concludes that five of nine voucher programs saved the state money, and the remaining four had a fiscally neutral effect on the state budget. All of the tax-credit programs

considered produced costs for the state, but yielded higher savings for the public school districts. With respect to equity, school choice programs seem successful in reducing inequalities in the educational system. For example, Ladner and Burke (2010) show that the reforms in Florida have been closing the achievement gap for Hispanic students. We must also consider the impact of educational reform on the existing public system. In Milwaukee, public schools that faced voucher competition made greater academic gains than similar schools that did not face such competition. Hoxby (2001) proves that by subject, public schools under competition made gains greater than control schools by 3 percentage points per year in math, 5 points per year in science, 3 points per year in social studies, and 3 points per year in language. Greene and Forster (2002) found that schools with 100% of the student body eligible for vouchers made academic improvements 15 percentage points higher than schools with only 50% of the student body eligible for vouchers. Similar results were found in Florida schools facing the threat of vouchers (West and Peterson, 2006). The positive effect of school choice on public schools also extends on a larger scale to schools near other towns offering school choice programs. This was found in Maine and Vermont, in which public schools located near towns that began offering school voucher programs began to see academic improvements 12% higher than before the programs were established (Hammons, 2002).

There is no simple method for measuring parental satisfaction. Opinion surveys would allow us to assign some quantity to satisfaction, though it seems unreasonable to expect parents to identify to what degree their satisfaction has increased or decreased as a direct result of school choice. A more sensible proxy is to consider whether a parent's participation in school functions increases or decreases after implementation of school choice. Gleason (2010) found that parents of lottery winners that were offered admission to study at charter schools were significantly more likely to volunteer or attend school activities than parents of lottery losers. In contrast, the study found that parents of lottery losers were significantly more likely to be members of the parent-teacher association than would parents of lottery winners. It is expected that parents who are more satisfied with their children's schooling would be more likely to participate in voluntary roles, while those who are dissatisfied would participate in groups such as the "Parent-Teacher Association" to be more active in influencing school policy.

Another very important dimension one should take into account is what will happen with dropout rates under school choice programs. The problem starts at measurement: Forster (2007) mentions that school systems have different ways to account for this statistic, and even if one is to believe the results, it is often difficult to disseminate the reasons why students dropout. More research is necessary, but early work by Greene (2004) addresses the issue. He finds a 36% dropout rate from the Milwaukee voucher program, significantly lower than the 59% found under Milwaukee's selective public high schools, or the 64% for the Milwaukee's public schools as a whole. Lastly, researchers can examine the effects of freedom of choice on transmission of civic values in the classroom and social integration. Some critics of school choice believe that allowing parents and students greater choice in schooling will lead to higher segregation. Greene

(1998), based on evidence from the National Education Longitudinal Study (NELS) argues the opposite may be true.

CONCLUSION

We hope to provide a solid introduction to school choice analysis. We present an overview of school choice, a discussion of the tools economists use to analyze school choice, and an economic perspective on school choice research thus far from both a theoretical and empirical perspective.

Because economists typically have a difficult time persuading municipal school districts or state governments to undergo social experiments, existing analysis depends heavily on what policy reforms are available. The popularity of voucher programs indicates that these are the easiest or most likely reforms to pass (probably because they satisfy voters' desires for improved equity in education). Tax credits are less popular, though some programs have been around long enough for more thorough research to take place. Schaeffer (2007) suggests that tax credits may become more popular in the future because they are less likely to be challenged in court, less likely to find united opposition, and higher-income individuals are more likely to support them over vouchers.

Experiments typically establish a randomized system of rewarding public financing, thus allowing for a control group. This method should not change in future reforms given that random selection is a principal component of statistical inference. Results from the literature indicate that school choice programs have been able to increase student achievement and school quality while reducing inequalities in the educational system without additional costs on state governments. There also seems to be a positive external effect on the existing public schools that face voucher competition. Finally, there is increased parental participation as well as a significant fall in dropout rates. Though results are encouraging, researchers call for caution in the interpretation of these findings since the ideal evaluation of these programs require more time to go by. One additional area of research that deserves increase attention concerns *how* information is presented to parents and *how* parents use this information to make decisions. Greene et al. (2010) show that the form in which information is presented to parents has important effects on their choice of school, and emphasize that all information presented should be characterized as relevant and easy to comprehend. Poor presentation of information may inhibit students and schools from reaching market equilibrium, and will understate the influence of schooling reform.

ENDNOTES

[1] Address: One Trinity Place; Department of Economics; Trinity University; San Antonio, Texas 78212-7200. Phone: (210) 999 – 8362. Email: rsantos@trinity.edu.

[2] A famous example is the "Texas Miracle;" Houston School Superintendent Rod Paige made his principals and administrators accountable for the performance of their students. Given hundreds of students per year

dropped out of Houston schools, principals and administrators responded to Paige's policy by reclassifying dropouts to improve their schools' numbers. In 2001, 463 students dropped out of Sharpstown High School, though none of them were listed as dropping out – every student either "transferred schools," "returned to their home country," or some other acceptable excuse. Please see http://www.cbsnews.com/stories/2004/01/06/60II/main591676.shtml for additional information.

[3] Opposing this condition, some programs, like the one in Washington D.C. were underfunded and required a tedious application process.

[4] The Georgia Special Needs Scholarship allows some students with special needs to transfer to another public school, public district, participating private school, or state school. Scholarships average about $6,000.

[5] The John M. McKay Scholarships for Students with Disabilities Program allows parents of children with special needs to change public schools if a parent is unsatisfied with their child's public school. The average scholarship in recent years has been over $7,000.

[6] Notice that this complaint also presumes that choosing a school for nonacademic reasons is inherently counterproductive. While such an act may ignore the primary objectives of school choice reform (that is, increasing school quality, increasing efficiency in the production of education, creating greater equity), it still may be the case that this choice leads to higher student accomplishment. For example, Coleman et al. (1982) argue that students who choose a school for religious reasons would form a supporting community of students and parents that share a common faith that generates social capital in the form of better networks for educationally productive relationships.

[7] Forster (2007) provides an interesting summary of these studies in table 2, page 39.

[8] The NCE, Normal Curve Equivalent, is a way of standardizing student test scores. A summary of the NCE system can be found at http://www.rochesterschools.com/Webmaster/StaffHelp/rdgstudy/nce.html.

REFERENCES

Aud, S. (2007). *Education By the Numbers: The Fiscal Effect of School*. Milton and Rose D. Friedman Foundation, School Choice Issues.

Barnard, J., Constantine, F., Jennifer, H., Donald, R. (2003). Principal Stratification Approach to Broken Randomized Experiments: A Cast Study of School Choice Vouchers in New York City. *Journal of the American Statistical Association*. Vol. 98, No. 462, 318-320.

Campbell, D.T., and J.C. Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.

Chubb, J.E., and T.N. Moe (1990). *Politics, Markets, and America's Schools*. Washington, D.C.: Brookings Institution Press.

Coleman, J, T Hoffer, and S Kilgore (1982). Cognitive Outcomes in Public and Private Schools. *Sociology of Education.* Vol. 55, No. 2/3, 65-76.

Coons, J.E., and S.D. Sugarman (1978). *Education by Choice: The Case for Family Control*. Berkeley, CA: University of California Press.

Coulson, A. (2008). The Fiscal Impact of a Large-Scale Education Tax Credit Program. *Policy Analysis*. 618.

Cowen, J. (2007). School Choice as a Latent Variable: Estimating the "Complier Average Causual Effect" of Vouchers in Charlotte. *Policy Studies Journal*, vol.26 No2, 301-315.

Cullen, J., Jacob,B., and Levitt, S. (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica,* 1191-1230.

Deming, D. (2009). *Better Schools, Less Crime?*. Working Paper.

Epple, D., Richard, R. (2002). *Educational Vouchers and Cream Skimming*. National Bureau of Economic Research, Working Paper No 9354.

Fantini, M.D. (1973). *Public Schools of Choice*. New York: Simon and Schuster.

Finnigan, K., Tricia, S. (2009). *Interdistrict Choice as a Policy Solution: Examining Rochester's Urban-Suburban Interdistrict Transfer Program (USITP)*. National Center on School Choice.

Fiske, Edward, and Helen Ladd (2000). *When Schools Compete: A Cautionary Tale*. Washington, D.C.: Brookings Institution Press.

Forster, G. (2005). *Using School Choice: Analyzing How Parents Access Educational Freedom*. Milton and Rose D. Friedman Foundation, School Choice Issues.

Forster, G. (2007). *Monopoly vs. Markets: The Empirical Evidence on Private Schools and School Choice*. Milton and Rose D. Friedman Foundation.

Forster, G. (2009). *A Win-Win Solution: The Empirical Evidence on How Vouchers Affect Public Schools*.Milton and Rose D. Friedman Foundation, School Choice Issues.

Forster, G., D'Andrea, C. (2009). *An Empirical Evaluation of the Florida Tax Credit Scholarship Program*. Milton and Rose D. Friedman Foundation, School Choice Issues.

Friedman, M. (1955). *The Role of Government in Education*. Economics in the Public Interest. R.A. Solo. New Brunswick: Rutgers University Press.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.

Gleason, P. (2010). *The Evaluation of Charter School Impacts*. National Center for Education Evaluation and Regional Assistance. U.S. Department of Education, NCEE 2010-4029.

Greene, J. (1998). *Civic Values in Public and Private Schools*. Learning From School Choice. Paul E. Peterson and Bryan C. Hassel. Washington D.C.: Brookings Institute.

Greene, J., Peterson, P., Jiangtao, D. (1998). *School Choice in Milwaukee: A Randomized Experiment*. Learning From School Choice. Paul E. Peterson and Bryan C. Hassel. Washington D.C.: Brookings Institute.

Greene, J., Greg, F. (2002). *Rising to the Challenge: The Effect of School Choice on Public Schools in Milwaukee and San Antonio*. Civic Bulletin, No. 27.

Greene, J. (2004). *Graduation Rates for Choice and Public School Students in Milwaukee*. School Choice Wisconsin.

Greene, J., Loveless, T., MacLeod, W., Nechyba, T., Peterson, P., Rosenthal, M., Whitehurst, G. (2010). *Expanding Choice in Elementary and Secondary Education*. Brown Center on Education Policy, Brookings Institution Press.

Hammons, C. (2002). *The Effects of Town Tuitioning in Maine and Vermont*. Milton and Rose D. Friedman Foundation, School Choice Issues.

Hanushek, E., Klain, J., Rivkin, S. (2001). *Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools*. National Bureau of Economic Research, Working Paper No. 8479.

Howell, W., Peterson, P. (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, D.C, Brookings Institution Press.

Hoxby, C. (2000). Does Competition Among Public Schools Benefit Students and Taxpayers?.*American Economic Review.* Vol. 90, No. 5, 1209-1238.

Hoxby, C. (2001). Rising Tide. *Education Next.*

Jencks, C. (1966). Is the Public School Obsolete?. *Public Interest.* 18-27.

Ladner, M., Linsey, B. (2010). *Closing the Racial Achievement Gap: Learning from Florida's Reforms*. Backgrounder, No. 2468.

Leung, R. (2004). *The 'Texas Miracle*. CBS News 25 Aug 24.

MacLeod, W. B., Urquiola, M. (2009). *Anti-Lemons: School Reputation and Educational Quality*. National Bureau of Economic Research, Working Paper, No. 15112.

McCluskey, N. (2005). Corruption in the Public Schools: The Market is the Answer. *Policy Analysis*, No. 542.

Merrifield, J. (2008). Dismal Science: The Shortcomings of U.S. School Choice Research and How to Address Them. *Policy Analysis,* No. 616.

Peske, H., Haycock, K. (2006). Teaching Inequality: How Poor and Minority Students are Shortchanged on Teacher Inequality. *Education Trust.*

Peterson, P. (1990). Monopoly and Competition in American Education. Choice and Control in American Education, Vol 1: *The Theory of Choice and Control in Education.* W. H. Clune and J.F. Witte. New York: Falmer Press.

Peterson, P. (1999). *Top Ten Questions Asked About School Choice.* Brookings Papers on Education Policy, No.2, 371-418.

Salisbury, D. (2005). Saving Money and Improving Education: How School Choice Can Help States Reduce Education Costs. *Policy Analysis,* No. 551.

Schaeffer, A. (2007). The Public Education Tax Credit. *Policy Analysis,* No. 605.

U.S. Department of Education, National Center for Education Evaluation and Regional Assistance (2010). *The evaluation of charter school impacts,* NCEE 2010-4029.

Wells, A.S. (1993). *Time to Choose: America at the Crossroads of School Choice Policy*. New York: Hill and Wang.

West, M., Peterson, P. (2006). The Efficacy of Choice Treats within School Accountability Systems: Results from Legislatively Induced Experiments. *The Economic Journal.*

Wolf, P., Gutmann, B., Puma, M., Rizzo, L., Eissa, N. (2007). *Evaluation of the D.C. Opportunity Scholarship Program: Impacts after One Year.* U.S. Department of Education, NCEE 2007-4009.

FOREIGN AID AND DEVELOPMENT: WHAT CAN DEVELOPING NATIONS LEARN

Tamer Rady, Ain-Shams University

ABSTRACT

In the wake of the Arab spring revolutions voices within that region started to question foreign aid. While the loud voices are clear in the media, none provided a scientific rational of the effect of aid. This paper reviews the recent findings on the relationship between foreign aid and development. Findings indicate that despite conflicting empirical results on the nature of relationship that exists between aid and growth, foreign aid can achieve some development goals in absence of corruption, with good governance present in the recipient country, with the presence of sound economic policy and institutions. Findings also indicate that aid's impact may be larger the more donors coordinate among themselves. In addition, aid in the form of technical assistance can at times be of larger impact.

INTRODUCTION

Since the revolutions and regime changes took place in Egypt and Tunisia, offers of foreign aid made headline news. At the same time, some voices within the region started to renounce aid. In Egypt, an independent campaign called for Egyptian donations to match the sum of the annual U.S. aid Egypt is receiving. None of the objecting voices gave any scientific explanation for doubting the value of aid. This paper reviews recent literature on the relationship between aid and growth to unravel recent findings in the hope of making some solid recommendations about how aid should be viewed.

WHAT IS FOREIGN AID

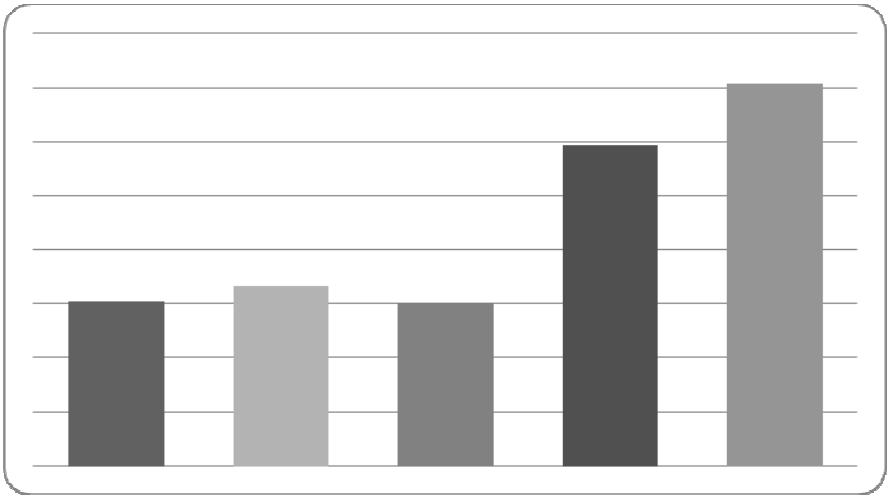
There is no one agreed upon definition of foreign aid. The American Heritage Dictionary defines foreign aid as economic or military assistance offered by one nation to another. The Britanica Concise encyclopedia defines aid as transfer of capital, goods, or services from one country to another. Lancaster (2006) defines foreign aid as a voluntary transfer of public resources from a government to another government, NGO, or an international organization with at least 25% grant element with the objective of improving the human conditions in the country receiving the aid. The last definition is similar to that given by the Development Assistance committee of the Organization for Economic Cooperation and Development without the

inclusion of the condition that the transfers are made from high income to low income countries, and including improving human conditions rather than the term development.

Foreign aid has many forms; it can be given in the form of capital transfers or in the form of technical assistance in consultation and training for civilian or military purposes. There are many reasons why governments give aid. Some aid is given for political reasons such as preserving peace. An example of that is the case of Egypt and Israel where the U.S. Secretary of State will have to annually declare that Egypt is committed and observing the peace treaty in order for the Congress to approve the yearly aid which is directed to military and civilian purposes. Other is given for specific development purposes whether increasing the capacity of certain industries, reducing poverty, or improving education or health care. Some is given to improve terms of trade and some can be given for humanitarian relief efforts in certain cases of emergency such as famine and natural disasters. While the political side of aid in many cases may be hard to investigate, the effect of aid in fulfilling its declared objective can be searched.

Prior to World War II the use of aid was rather scarce. However, it became obvious after the war that the United States should interfere to rebuild Europe, the political and economic dimensions of aid were apparent; a stable and productive Western Europe can stand against the spreading communist ideas and the Soviet efforts in this regard. Hogan (1987) provides a detailed description of the Marshall plan to rebuild Europe; $12 billion were funneled to 16 Western European nations to rebuild their economies in addition to fortify liberal economic policies. Through the years that followed, foreign aid became a popular policy tool of developed nations and its amounts increased dramatically. Figure one indicates that within the last twenty years, the total value of annual development assistance given increased from $60 billion to over $140 billion.

Figure 1: Official Development Assistance in Billions of $U.S.



Source: Compiled from OECD Aid Database

FOREIGN AID AND ECONOMIC GROWTH

Since foreign aid can be viewed as foreign capital early theories to explain its effect on growth resorted to traditional economic development theories and the role of capital stock and flows in development. Much of the early empirical research utilized framework offered by development economists such as Harod (1939), Domar (1946), Rostow (1960) and other variations of those models. The majority of work done assigns growth rate as the independent variable and aid as one of the dependent variables in addition to the traditional variables assigned in different development models and run regressions to evaluate the significance of the relationship between aid and growth. Research work from the 1970's, 80s and 90s have indicated conflicting results that varied from aid being positively linked to growth to negatively linked or insignificant. Griffin and Enos (1970) analyzed data from 12 countries that covered the years 1957-1964 and regressed average growth rate of GNP to the ratio of GNP to foreign aid to come up with a negative significance. In their attempt to explain it, they analyzed data from 32 developing nations between 1962 and 1964 and concluded that gross domestic savings as a percentage of GNP is inversely related to foreign savings as a percentage of GNP. That is foreign aid hinders domestic savings and hence negatively affects growth. Over (1975) expanded the data set and assumed foreign aid to be endogenous and came to a conclusion that is the opposite of Griffin and Enos; aid is positively linked to growth and can increase domestic savings. Mosley et al. (1987) argued that augmenting foreign aid into capital in the frame work of a simple Harrod-Domar model can be misleading. They argued that the impact of aid is not straightforward and that there are effects at three different levels to be considered; the direct effect aid allocation, the indirect effect on the spending pattern of the public sector of the recipient country, and the effect on the whole of government policy. They utilize a model that incorporates the three effects and simulates public sector behavior in a developing country where the government tries to use aid money to maximize its own welfare faced by the budget constraint and can take many actions of what to do with aid funds such as tax reduction, increasing productive and/or non-productive expenditures and others. Their empirical findings after analyzing data that cover from the late 1950's to the early 1980's for many developing countries indicate a non-significant relationship between aid and growth.

The 1990's have witnessed more research trying to explain the conflicting findings on the nature of the relationship between aid and growth. Bowen (1995) investigated both the direct and indirect effect of foreign aid on growth. Bowen Started from an initial model of growth where the annual GNP growth rate depends on domestic savings, foreign capital inflows, and annual growth rate of exports and expanded the model to incorporate the variations in the relationship between economic growth and its determinants according to the level of aid by adding aid to the parameters of the independent variables. Using a data base that covers 67 less developed countries between the period 1970 and 1988 Bowen's findings indicate that the aid-growth relationship is dramatically different across levels of development. This finding may

explain the non-significant results on the aid-growth relationship when investigated using linear models; the negative and positive relationships at different income levels may cancel each other out. Bowen's findings indicate that aid has been given in response to recipient need, as indicated by per capita income level. Findings point out that in countries with a per capita income of less than $987 aid was substituted for domestic savings and those nations had lower economic growth where in countries with a per capita income level above $987, aid was positively correlated with growth. The positive correlation may have been because recipient governments reduced their switching behavior as income rose and hence allowing foreign aid to act as a complement to domestic capital formation instead of a substitute, or may have been because the effectiveness of other determinants of growth such as domestic investment or exports increased with the level of development.

Boone (1996) tried to relate the effectiveness of foreign aid programs to the political regime of recipient countries. He presented a model where poverty is caused or increased by distortionary policies introduced by politicians. His model is demonstrated a maximization problem where politicians want to maximize the use of distortionary income tax and foreign aid in financing public spending and transfers and consumers want to maximize their utility. The model was tested on data that covers around 97 developing nations from the period of 1971 to 1990. Boone's findings indicate that aid increased consumption and that higher consumption did not benefit the poor. His findings also indicate that aid had an insignificant impact on improvements in basic measures of human development such as infant mortality and primary schooling ratios. Boone's paper introduced political determinants of aid as a factor in explaining the aid-growth relationship. However, an argument can be made that some governments may be increasing consumption to dampen the effect of the economic reforms path that could have been chosen.

The World Bank (1998) issued a report in which it stated that aid is more effective when given to nations with sound policies. In addition, the report indicated that beyond a certain level, aid can cause negative effect. The involvement of the domestic economic policies role in using aid gave way to more research in this area. Burnisde and Dollar (2000) investigated whether the effect of aid on growth is conditional on economic policies and whether donor governments and agencies allocate more aid to countries with good economic policies. Using a neoclassical growth model Burnside and dollar allowed for economic policy variables to interact with aid. The model treats a gift of aid as an additional capital that should be positively linked with growth, however, if economic policies affected growth negatively then so would the aid. To define good policies Burnside and Dollar chose from the economic variables that are associated with growth; the budget surplus, the inflation rate, and index of openness of a country that was developed by Sachs and Warner (1995). They tested their model on data for 56 countries divided into six four- year time periods from 1970-1973 until 1990-1993. Their findings indicate that aid has a positive impact on growth in developing countries with good fiscal, monetary, and trade policies and little effect in the presence of poor policies. Good policies are ones that are

themselves important for growth. In addition, their findings suggest that the quality of policy in the recipient country has only a small impact on the allocation of aid. Findings imply that aid would be more effective if it were conditioned on good economic policy. Another interesting finding of their work is that aid which is multilaterally managed was allocated in favor of good policy, where bilateral aid had a strong positive correlation with government consumption. That finding may explain why the impact of aid on growth may not be positive in some cases. The findings of Burnside and dollar spurred a lot of research that followed suit. Hansen and Tarp (2001) used similar model specification with the addition of more variables. They investigated the relationship between foreign aid and real GDP per capita growth as it emerges from simple augmentations of popular cross-country growth specifications. Their findings indicate that aid in all likelihood increases the growth rate, and contrary to the Burnside and Dollar the result is not conditional on good policy. They demonstrate that there exists a decreasing returns to aid, and that the estimated effectiveness of aid is very sensitive to the choice of estimator and the set of control variables. They point out that aid impacts growth via investment and hence when investment and human capital are controlled for they find no positive effect of aid. Dalgaard and Hansen (2001) used the same data set as Burnside and Dollar. They developed a neoclassical growth model where aid leads to growth even in economies where aid is consumed and demonstrated that in such model the interplay between good economic policy and aid is ambiguous. They argued that good policy may tend to reduce the growth effect of aid because they act as substitutes in the growth process. Their findings indicate that a more positive impact of aid on growth in good policy environments is not robust. They conclude that it is premature to apply policy selectivity rules in aid allocation and point out that applying the policy selectivity rule will increase returns to aid when returns are measured as the correlation between aid and growth in income per capita because good policy leads to higher growth.

More extensions of the Burnside and Dollar were introduced. Guillamont and Chauvet (2001) proposed that aid effects on growth are not necessarily positive but depend on specific conditions in each recipient country. They argued that aid effectiveness also depends on exogenous and mostly external environment factors such as the terms of trade trend, the real value of exports instability, climatic shocks, and others. Their findings indicate that aid seems to have accelerated growth only in the more vulnerable countries. That is, aid significantly dampened the negative effects of a bad environment or external shocks. They were not able to find that aid effectiveness (in growth terms) has been increased by a better government policy. They suggest that aid should be allocated to countries affected by poor environment or external shocks and also in countries where support for good policies are needed. Similar results were obtained by Collier and Dehn (2001) who incorporated export price shocks into the analysis of the effect of aid on growth. They introduced these large shocks into the Burnside and Dollar setup. Their findings indicate that the shocks are highly significant in reducing growth and as these shocks are included, the Burnside-Dollar results turns robust to choice of sample. They pointed out that the effects of negative shocks on growth can be decreased through increases in

aid. They conclude that targeting aid to countries experiencing negative shocks may be more important for aid effectiveness and reduction of poverty than targeting aid to countries with good policies; however their findings show that donor countries usually do not follow this policy. Lensink and White (2001) challenged the World Bank (1998) report that aid can have negative returns after a certain threshold. Their regression set up followed that of Burnside and Dollar (2000) and included variables that reflect good policy. Their data pool had 138 countries covering the period between 1975 and 1995. Their findings showed some support for the argument that aid may have negative returns at a certain high level of inflows beyond that predicted by the World Bank. Their findings also point out that both inflation and terms of trade as signs of good policy were significantly correlated with the use of aid funds.

Feng (2001) indicated that oppressive regimes in developing nations may deter private investment. Isham et al. (1997) findings point out that oppressive regime can cause a decrease in the return on public investment. Wright (2008) noted that two third of aid recipient countries during the period 1960 - 2002 were authoritarian regimes. However, he observed that some of those regimes managed to use aid well such as Botswana and South Korea between 1961 and 1987, where for the same period Zaire and Kenya did not use it effectively. Wright focused on how a dictator's time horizon formulates his incentives over how to use aid. The time horizon reflects whether the authoritarian regime is stable and thinks it has a higher probability to stay in power or have a higher probability of losing power due to perceived threats. Wright argued that the predicted probability of authoritarian regime failure serves as the best proxy for measuring time horizons. Wrigh's logic is that authoritarian regimes use of aid can be affected by the time they perceive to last in power in three ways; the first affects all types of regimes and is the incentive to invest in public goods and the protection of property rights and so on if they perceive that they will hang on to power for a long time. However, if they perceive higher threats to their seats they tend to consume state resources in ways that can harm economic growth by resorting to oppression and paying off their political opponents, which in turn directs resources to consumption rather than investment and increase rent seeking behavior. The last incentive when they are facing a short time horizon is to secure personal wealth in fear of what to do after they lose power. Analyzing data from1961 to 2001, and using proxies for political risk from International Country Risk Guide and democracy measures based on the work of Przeworski et al. (2000) and Cheibub and Gandhi (2004), Wright concludes that in authoritarian regimes with a low probability of regime failure there is a robust positive relationship between aid and growth and the opposite with dictators with higher probabilities of losing power.

Rajan and Subramanian (2008) recognized that the flow of aid is influenced by a country's situation as it may go to a country experiencing a natural disaster which may show a negative correlation between aid and growth, or it can flow to a country that effectively used aid in the past, and since growth exists already there may be a positive correlation between aid and growth. They went to investigate and under one framework, the robustness of the relationship of aid and growth across time horizons; medium and long run, and over periods; 1960s through

1990s, the sources of aid; multilateral or bilateral, types of aid; economic, social, food, etc., the timing of impact of aid; short-term or long-term, data specifications; cross-section or panel, and samples chosen to investigate. Their control variables included initial income, initial life expectancy, a measure of institutional quality, a variable indicating quality of policies, a measure of geography that reflects weather and land fertility, an indicator of external shocks that captures average growth and the variability of a country's terms of trade, government consumption, and index of revolutions and other variables. Their findings show little evidence of a robust positive impact of aid on growth. In their cross-sectional analysis they find some evidence for a negative but non-significant relationship between aid and growth in the long run. They also obtain evidence of a positive relationship for the period between1980-2000 when countries with extreme variables within the sample are included. In addition, they find no evidence that aid works better in better policy or institutional or geographical environments, or that certain types of aid work better than others.

Kurt and Stephen (2009) argue that the aid-growth literature is conflicting because it does not differentiate between foreign aid as technical assistance and non-technical assistance. They also argue that there exists economies of scale that is associated with technical assistance aid and that its effectiveness may depend on how the level of aid is fragmented. To theoretically and empirically support their argument they construct a neoclassical growth model frame work, where technical assistance induces knowledge spillovers that improve the productive capabilities in the recipient country through human capital. This spillover is affected by the degree of fragmentation of technical aid. On the other side, non technical assistance is treated as a transfer of income that increases the available resources for consumption and investment. They empirically test their theoretical model and their findings point out that technical assistance was positively correlated with growth and the less the degree of fragmentation of aid donors the higher the impact on growth. The question of donor cooperation in giving aid was raised in the literature. Raechelle and Todd (2006) raised the question whether donors cooperatively fund foreign aid. They tested for the underlying allocation process for foreign aid commitments in its numerous forms, when these commitments are viewed as yielding public benefits. They theoretically derive a set of reduced-form equations for non-cooperative Nash-Cournot and cooperative Lindahl behavior in a setting where aid may provide purely public benefits to all potential donors along with donor-specific benefits. For countries that do not follow either behavior they tested alternative forms of bureaucratic-based decisions. They empirically test their model and conclude that there is no evidence of donor cooperation.

There is plenty of research in the literature that investigates the relationship between aid efficiency and good governance in the recipient countries. Winters (2010) reviewed different accountability relationships that exist in foreign aid projects among donors, governments, implementing agencies and end users. In addition, he reviewed existing empirical evidence in the literature that inquire the relationship between aid effectiveness, democracy and governance in the recipient countries. Winters points out to prevailing evidence that aid is more effective in

countries with superior accountability institutions. That is, in countries where governments are more democratic and the rule of law is upheld, foreign aid is more likely to lead to economic growth and individual foreign aid projects are more likely to witness higher rates of return. Winters points out that aid projects that incorporate end user participation as direct accountability mechanism are more likely to be achieve their goals. Bourguignon amd Sundberg (2007) point out that that it is not surprising that cross-country evidence on aid effectiveness is weak and argue that this does not imply that all aid is ineffective, nor does it imply that little is known about how to make aid more effective. They stress the need to link aid to the development outcomes expected from it. They argue that such link would demonstrate three relationships that can be easily explored and where much is known; how funds are transferred from donors to policy makers, how policy makers can translate the funds to working policies, and how the policies shape the outcome. They argue that an aid model with those three links well examined, with donors aligned around the development strategy of the recipient, and aid allocation system that is ruled by good governance and based on results that well monitored will achieve its development goals.

CONCLUSIONS

This paper reviewed existing literature on the nature of the relationship between aid and growth. Findings indicate that there still exists a large degree of conflicting results about the nature of this relationship. While there may be more research pointing out to a more positive but somewhat insignificant relationship, much of the reason of such conflict may pertain to the different model specifications used and the statistical noise surrounding the process. Despite conflicting empirical results on the nature of relationship that exists between aid and growth, there is relatively high consensus that foreign aid can achieve some development goals in absence of corruption, with sound economic policy and institutions and when given to more stable governments. There is also evidence that aid's impact may be larger the more donors coordinate among themselves. In addition, recent literature points out that aid in the form of technical assistance can at times be of larger impact than non-technical aid.

REFERENCES

Boone, P. 1996. Politics and the Effectiveness of Foreign Aid. *European Economic Review. 40*(2), 289-329.

Bourguignon, F. & Sundberg, M. (2007). Aid Effectiveness - Opening the Black Box. *American Economic Review. 97*(2), 316–321

Bowen, J. (1995). Foreign Aid and Economic Growth: An Empirical Analysis. *Geographical Analysis. 27*(3), 249-261.

Burnside, C. & Dollar, D. (2000). Aid, Policies, and Growth. *American Economic Review. 90*(4), 847-868.

Cheibub, J. & Gandhi, J. (2004). Classifying Political Regimes: A six-fold Measure of Democracies and Dictatorships. (Working Paper). New Haven, CT: Yale University, and New York: New York University.

Collier, P. & Dehn, J. (2001). Aid, Shocks, and Growth. Working Paper No. 2688, World Bank.

Dalgaard, C. & Hansen, H. (2001). On Aid, Growth and Good Policies. *Jourrnal of Development Studies. August, 37*(6), 17- 41.

Domar, E. D. (1946). Capital Expansion, Rate of Growth and Employment. *Econometrica 14,* 137-147.

Feng, Y. (2001). Political freedom, political instability, and policy uncertainty: A study of political institutions and private investment in developing countries. *International Studies Quarterly. 45*(2), 271-294

Griffin, K. & Enos, J.L. (1970). Foreign Assistance: Objectives and Consequences. *Economic Development and Cultural Change. 18,* 313-27.

Guillamont, P & Chauvet, L. (2001). Aid and Performance: A Reassessment. *Jounral of Development Studies. 37*(6), 66-92.

Hansen, H. & Tarp, F. (2001). Aid and Growth Regressions. *Journal of Development Economics. 64*(2), 547-570.

Harod, R. (1939). An Essay in Dynamic Theory. *Economic Journal 44,* 14-33.

Hogan, M. (1987). *The Marshall Plan: America, Britain, and the Reconstruction of Western Europe, 1947–1952.* New York: Cambridge University Press.

Isham, J., Kaufmann, D. & Pritchett, L. (1997). Civil liberties, Democracy, and the Performance of Government Projects. *World Bank Economic Review. 11*(2), 219-242.

Kurt A. & Stephen K. (2009). Foreign Aid, Donor Fragmentation, and Economic Growth. *The B.E. Journal of Macroeconomics. 9*(1), 1-30.

Lancaster, C. (2006). *Foreign Aid: Diplomacy, Development, Domestic Politics.* Chicago and London: The University of Chicago Press.

Lensink, R. & White, H. (2001). Are There Negative Returns to Aid? *Journal of Development Studies. 37*(6), 42-65.

Mosley, P. , Hudson, J. & Horrell, S. (1987). Aid, the Public Sector, and the Market in Less Developed Countries. *Economic Journal. 97,* 616-641.

Over, A. M. (1975). An Example of the Simultaneous Equations Problem: A Note on Foreign Assistance: Objectives and Consequences. *Economic Development and Cultural Change. 23*(4), 751-56.

Przeworski, A. , Alvarez, M., Cheibub, J., & Limongi, F. (2000). Democracy and Development. Cambridge, UK: Cambridge University Press.

Raechelle M. & Todd S. (2006). Do Donors Cooperatively Fund Foreign Aid? *The Review of International Organizations. 1*, 337-357.

Rajan, R. G. & Subramanian, A. (2008). Aid and Growth: What does the Cross-Country Evidence Really Show? *The Review of Economics and Statistics. 90*(4), 643–665.

Rostow, W. W. (1960). The Stages of Economic Growth: A Non- Communist Manifesto (Second Edition). London: Cambridge University Press.

Sachs, J. D. & Warner, A. M. (1995). Economic Reform and the Process of Global Integration. *Brookings Papers on Economic Activity. 1,*1-118.

Winters S. W. (2010). Accountability, Participation and Foreign Aid Effectiveness. *International Studies Review. 12,* 218–243.

World Bank (1998). *Assessing Aid: What Works, What Doesn't and Why.* New York, NY: Oxford University Press.

Wright, J. (2008). To Invest or Insure? How Authoritarian Time Horizons Impact Foreign Aid Effectiveness. *Comparative Political Studies. 41*(7), 971–1000.

WHAT FACTORS INFLUENCE FIRST-YEAR STUDENTS' SUCCESS AND FAILURE AT TRU: A CASE OF INTRODUCTORY ECONOMICS COURSES

Mohammad Mahbobi, Thompson Rivers University

ABSTRACT

The likelihood of success and failure in Microeconomics, ECON 1900, and Macroeconomics, ECON 1950, of 1369 Thompson Rivers University (TRU) students was analyzed using a probit model. This approach uses students' final marks in these courses along with other personal and institutional characteristics. It also determines the probability of passing the two introductory courses in economics against alternative contributing factors. The success and failure have found to be influenced by factors such as high school GPA, age, gender, location, type of high school, etc.. In addition, the results indicate that boys have a higher chance of successfully passing Microeconomics while such differentiation between boys and girls was not clearly evident in case of Macroeconomic. Prior academic achievement, measured by the average GPA of some selected high school courses, was also found to be significant. In addition, the likelihood is also dependent on the age of students at the time the course is taken. For a given high school GPA and regardless of student's gender, the highest probability of passing these courses occurs at age 26. Finally, students of both genders who were recruited from the School Districts close to Thompson Rivers University have shown better chances to pass these courses.

INTRODUCTION

The likelihood of success and failure in Principles of Microeconomics, ECON 1900, and Principles of Macroeconomics, ECON 1950, of 1369 TRU students were analyzed using a probit model. TRU students' final marks along with their personal and institutional characteristics were used to determine the probability of passing the two introductory courses in economics. Statistical analyses of the academic performance of university students have been widely used in many empirical studies including Anderson, Benjamin, and Fuss (1994), Smith and Naylor (2005), Birch and Miller (2006), Geide-Stevenson (2009), and Bethune (2010). In order to give a better picture of the performances of TRU first-year student in introductory economics courses, this paper utilizes a probit model in determining the probability of success and failure along with its influential determinants. The structure of this paper is as follows. First, main influences on first year TRU students' performances will be identified. Second, the estimation technique will be explained. Empirical results will then be presented, followed by conclusions and perspectives.

Objectives

The general objectives of this paper are to examine relationship between prior academic achievements in high school and probability of success and failure in the first year of university at TRU. The specific goals of this poster are to identify the determinants of success/failure, the impacts of institutional characteristics, and the influences of Personal Characteristics of TRU students in Introductory Economics Courses.

Model Specification and Data

The likelihood of success and failure in Microeconomics, ECON 1900, and Macroeconomics, ECON 1950, of 1369 TRU2[2] students during 2002 to 2009 was analyzed using a probit model. Students' final marks along with other personal and institutional characteristics are used to determine the probability of passing the two introductory courses in economics. Probability of failure in ECON 1900, and 1950 were applied to the cases when students received a W (withdraw), DNC (Did Not Complete), D (for 50), or F (less than 50). All these cases considered as failure. In the probit model the success and failure of students were related to determinants such as high school GPA, age, gender, possible gap between the time that student was graduated from high school and the time that the course was taken, location of high school, and type of high school. A dummy variable used for location of high school. All high schools within the School District 73 were grouped as TRU main recruitment region. The type of high school was complied according to British Columbia Education Ministry. The high school GPA was calculated based on the average of the available high school academic courses taken by the student. These include principles of mathematics 12 or application of mathematics 12, calculus12 (if taken), Physics 11, and 12, Chemistry 11 and 12, and English 11 and 12.

LITERATURE REVIEW

The issue of students' success and failures have long been in the centre of education of economics. With a focus on the factors that determine a student's success in introductory economics course, Anderson, Benjamin, and Fuss (1994) have applied ordinary least squares to estimate the student's success in introductory university economics courses. They have selected a total of 6718 students who were enrolled in ECO 100 in two campuses of University of Toronto during winter 1988 and summer 1989. The most important factors found to be significant are achievement level (grade 13 average), the taking of a course in calculus a high school senior students.

Birch and Miller (2006) paper on student outcomes at the Western University in Australia discusses how personal and high school characteristics, as well as university entrance scores

influence students' academic success. They have applied quantile regression to investigate how the effects of these factors vary along the grade distribution. Their research shows that the factors that influence grades have a more pronounced impact on the success of low-achieving students than that of high-achieving students. Factors that were used in their observations include personal characteristics, contemporary work patterns and prior academic achievement (PAA). Studies such as this have also been done in the United States, and these studies have all shown that there is a strong relationship between student outcomes at university and PAA.

Bethune (2010) explored the relationship between attendance and learning in the principles of microeconomics in a small liberal college with a focus on teaching excellence. By using a specific test on student knowledge of economics during the first day of class, and the final examination of the course, and with application of stepwise regression, Bethune found that absenteeism has a significant correlation with learning introductory economics. Further, the student absenteeism has shown strong relationship to course grade.

Geide-Stevenson (2009) has built a model of reduced form educational production function that was suggested earlier by Emerson and Taylor (2004). Students' achievements from Weber State University in the Utah state were formulated as a function of academic ability, student characteristics, number of absences, and membership in the experimental group. The experimental groups were set up according to Gorve and Wasswerman (2006). Geide-Stevenson found that more inexperienced junior students found to benefit from graded homework assignments or more individualized feedback.

EMPIRICAL RESULTS

This study analyzes TRU first year students' successes in two introductory courses in economics. Table-1 and Table-2 show the output of the estimated probit model for ECON 1900, ECON 1950 respectively. The GAP variable- the lag of time between the time students was graduated from high school and the time these courses were taken- was not found to be significant. In addition, gender was not significant in principles of macroeconomic. The latter simply indicates that while there is no difference among boys and girls in their probabilities of success or failure in taking principles of macroeconomic, gender of student is important (with 0.10 level of significance) in successfully passing or failing principles of microeconomic. Table-1 presents the estimated probit model for both courses.

One implication of the estimated coefficients of the probit model is to plot a probability response curve. The predicted probabilities may vary with an independent variable. One may put some important and related choices of variables together and predict such probabilities. For the probit model above, suppose we are interested in the effect of gender on the probability of success for boys and girls under specific set of assumptions. In this paper, the following scenarios are discussed.

Table-1 The Estimated Probit Model for both Introductory Economics Courses				
Determinants	Estimated Coefficient	Standard Error	Estimated Coefficient	Standard Error
Intercept	-3.668	1.155	-3.039	1.387
Location	0.1455	0.032	0.0732	0.034
High School GPA	0.4129	0.022	0.2567	0.022
Age	0.2831	0.108	0.2486	0.127
Age Square	-0.0055	0.002	-0.0047	0.0029*
Gender	-0.0554	0.030*	-	-
High School Type _PH	0.0523	0.032*	0.2159	0.0348
High School Type _IH	-02351	0.069	0.0232	0.0717*
Log likelihood	-4522.77		-4094.78	
Note: Dependent variable is probability of grades in ECON 1900 and ECON 1950 greater than 55 or C-. Sample size is 1369 students with different available high school GPA in different courses. * indicates significance at the 10% level. The rest of variables are significant at 5% level. High School Type _PH is an indicator of Public high school, and High School Type _IH stands for Independent High School.				

The first scenario is to divide the students into two groups with specific characteristics. Students who were recruited from within the TRU region, i.e. School District #73 and students arrived at TRU from the rest of the province and Canada[3].Under this scenario, the probability of successfully passing ECON 1900 for a high school student who graduated either from public or independent school, and was recruited from the S.D.#73 with an average GPA of high school courses is depicted in Figure-1.

As shown in Figure-1, age and such probability are positively correlated. The probability, however, sharply increases at age below 22 and starting to decline after age 26. In other words, the probability of successfully passing microeconomics is at maximum with a 26 years old student's regardless of the student's gender. However, the highest probability goes to the boys with 0.83. At every other level of age, girls with average high school GPA show slightly lower chance to pass this course. Figure-2 shows the probabilities of successfully passing ECON 1900 both for boys and girls who were recruited from the region other than the School District #73. This scenario the same as the first one except the probabilities were estimated for students out of the School District #73. The rest of the assumptions of the first scenario are still valid.

As it is clear, compared to Figure-1 such probabilities are is generally lower for students who were recruited from a region outside School District #73. The predicted declining point for the probability is still at age 26.

Figure-1 Probability of Successfully Passing ECON 1900-Scenario#1A

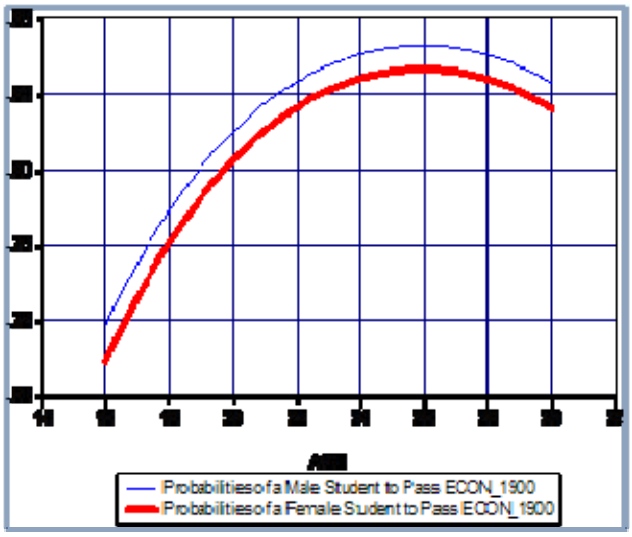
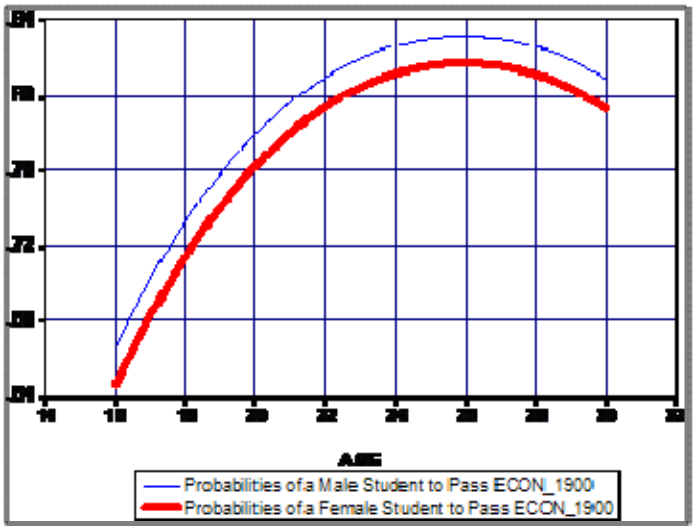


Figure-2 Probability of Successfully Passing ECON 1900-Scenario#1B



In order to examine the impact of high school GPA on the probability of passing ECON 1900, we select students with 26 at which we expect the highest possible success in principles of Microeconomics. Figure-3 and Figure-4 present cases where the estimated probabilities are

depicted against different levels of high school GPA for such students. This has been estimated for both gender and region and named as scenario#2A and scenario#2B.

Figure-3 Probability of Successfully Passing ECON 1900-Scenario#2A

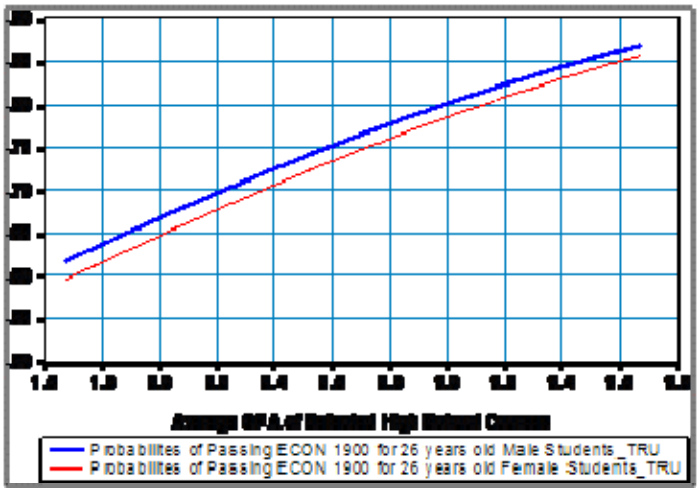
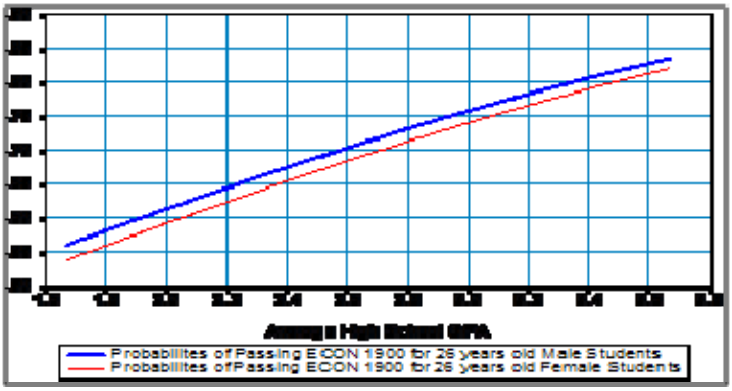


Figure-4 Probability of Successfully Passing ECON 1900-Scenario#2B



The following two graphs show the probabilities of successfully passing ECON 1900 for a typical A+ and C- students who was recruited from within the close vicinity of TRU. The two related scenarios are named Scenario_3A and Scenario_3B. Figurre-5 and Figure-6 represent such cases for boys and girls respectively.

Figure-5 Probability of Successfully Passing ECON 1900-Scenario#3A

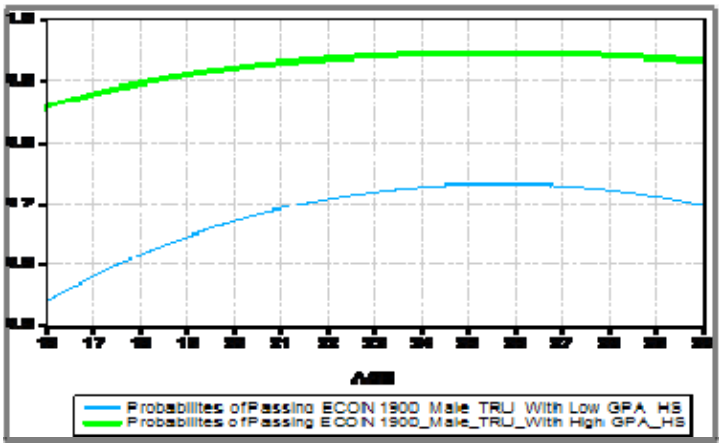
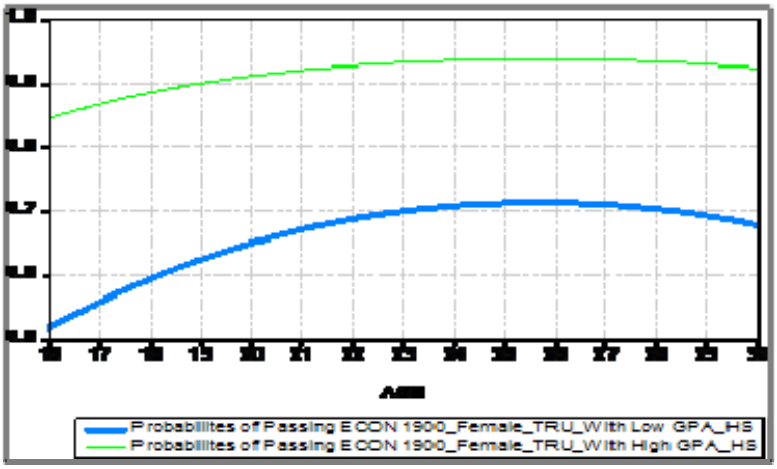


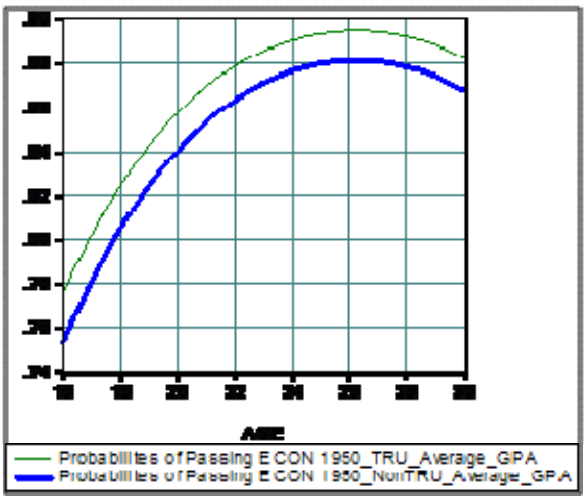
Figure-6 Probability of Successfully Passing ECON 1900-Scenario#3B



For principles of macroeconomic course, we may also want to estimate impacts of all these factors under different scenarios. For the probit model above, we now turn to the effect of students who recruited from within the TRU vicinity and those who have come to TRU from other regions. Since gender was not found to be significant for principles of macroeconomic, the probability of success for TRU and non-TRU high school students under alternative set of assumptions will be explored. Scenario#4 is designed for students with average GPA who have been graduated from either public or independent high schools. Figure-7 represents the

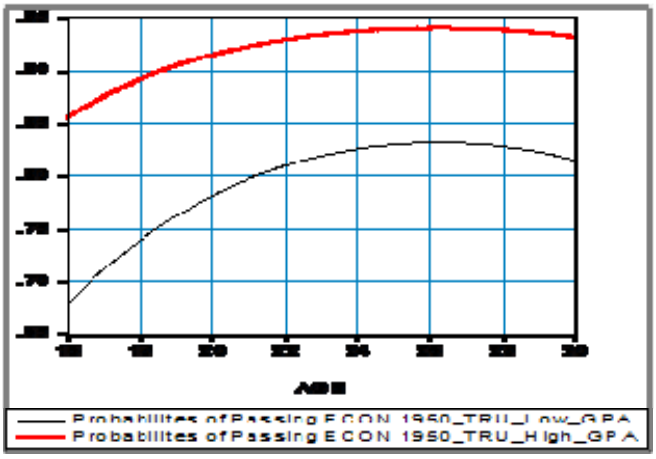
probabilities of successfully passing ECON 1950 for TRU versus non-TRU regions over different range of age for students with an average high school GPA.

Figure-7 Probability of Successfully Passing ECON 1950-Scenario#4



The result indicates that students with an average high school GPA, who were recruited from a region close to TRU, have shown higher probabilities to successfully pass principles of macroeconomics. Such likelihoods can also be estimated for students who had graduated from public and independent high schools close to TRU region. We define Scenario#5 by dividing these students within two groups with low and high GPA. Low high school GPA was set 1.66 and high GPA was assumed to be 4:00. Figure-8 represents the estimated likelihoods for such students.

Figure-8 Probability of Successfully Passing ECON 1950-Scenario#5



Interesting fact here is that students with high GPA who directly comes from a public or independent high school with no gap and with high GPA show very good chance to pass ECON 195, and this chance will stable over higher range of age. Students with low high school GPA, however, will have only 67% chance to successfully pass this course. Older students with low GPA will have the chance to increase their likelihood of passing ECON 1950as high as 84% at age 26. Figure-9 shows the estimated probabilities under scenario#6. Under this scenario the probabilities were estimated for students with the same set of characteristics as in scenarios#5 except they were recruited from a region other than School District #73.

Figure-9 Probability of Successfully Passing ECON 1950-Scenario#6

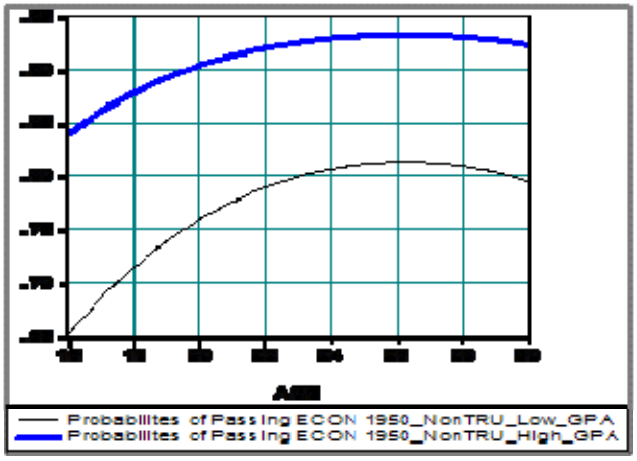
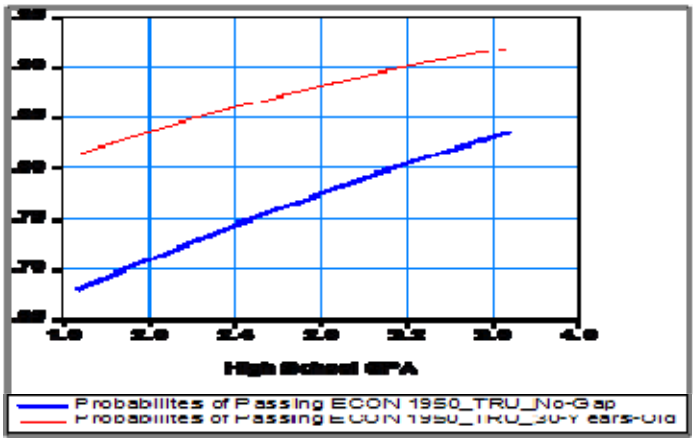


Figure-10 Probability of Successfully Passing ECON 1950-Scenario#7



The final scenario is to estimate the likelihoods of passing principles of macroeconomic for a freshman and a mature student who both recruited from within the TRU region. Figure-10 represents these probabilities for the two groups of students against their high school GPAs. As the graph indicates, a more mature student may pass this course with more than 82% chance even with low GPA. Such chance for a freshman is 63%. Also, as their GPA increases, the difference between the two likelihoods shrinks, indicating that to some degree age and GPA may be considered as complements to each other in passing this course.

CONCLUSIONS

This study utilized a probit model to examine success and failure in two introductory courses in Economics during the first year of study at Thompson Rivers University of Canada for the 2002 to 2009 entrance cohorts. Age, location of the student's high school, previous academic performances measured by the GPA of selected high school courses, and type of high school were found significant in the estimated probit model. Student's gender was only found to be important in principles of microeconomic course. The estimated probit models for both principles of microeconomic and macroeconomic courses were then used to predict the likelihoods of success and failure in these courses under ten different scenarios. The results show that student's average GPA in selected high school courses, being recruited from the region close to TRU, and with no gap after their the high schools are among the most importance determinants for those students most prone to failure in both courses. The results also indicate that male students have a higher chance of successfully passing principles of microeconomics while such differentiation between boys and girls was not clearly evident in principles of macroeconomic. In addition, the likelihoods of success in both courses are also dependent on the age of students at the time the course is taken. For a given high school GPA and regardless of student's gender, the highest probability of passing these courses occurs at age 26. Finally, students of both genders who were recruited from the School Districts close to TRU have shown better chances to successfully pass these courses.

ENDNOTES

[1] Dr. Mahbobi is an instructor at the Department of Economics, Thompson Rivers University, 900 McGill Road, Kamloops, BC, Canada, V2C 0C8. The contact email address is mmahbobi@tru.ca.

[2] Data were collected from Thompson Rivers University Institutional Planning and Analysis Office.

[3] The region of the students has been divided into S.D. #73 and others. This includes all students who are either from the rest of BC, Canada, or from other countries.

REFERENCES

Anderson, G. Benjamin D., and Fuss M., 1994, The Determinants of Success in University Introductory Economics Courses. *The Journal of Economic Education.* Vol. 25, No.2, pp. 1-17.

Bethune J.J., 2010, Attendance, Grades, and Learning in the Microeconomic Principles Class, *Journal of Economics and Economics Education Research*, (11), No. 1, pp 53-64.

Birch, E. and Miller, P. 2006, Student Outcomes At University in Australia: A Quantile Regression Approach, *Australian Economic Papers,* Vol. 45, No. 1, pp. 1-16.

Emerson, T. L.N. and Taylor, B. A., 2004, Comparing Students Achievement Across Experimental and Lecture-Oriented Section of a Principles of Microeconomics Course. *Southern Economic Journal*, Vol. 70, No. 3, pp. 672-693.

Geide-Stevenson, D., 2009, Does Collecting and Grading homework Assignments Impact Student Achievement in An Introductory Economics Course? *Journal of Economics and Economics Education Research*, (10), No. 3, pp 1-14.

Gorve, W. A. and Wasswerman, T., 2006, Incentives and student Learning: A Natural Experiment with Economic Problem sets, *American Economic Review*, Vol. 96, No. 2, pp. 447-452.

Hill, Griffiths, and Lim, 2008, *Principles of Econometrics,* Third Edition, John Wiley and Sons, Inc. USA.

Smith, J., and Naylor, R. 2005, Schooling Effects on Subsequent University Performance: Evidence for the UK University Population, *Economics of Education Review*, Vol. 24, No. 5, pp. 549-562.

Wooldridge J.M., 2002, *Econometrics Analysis of Cross Section and Panel Data,* First Edition, The MIT Press, USA.