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

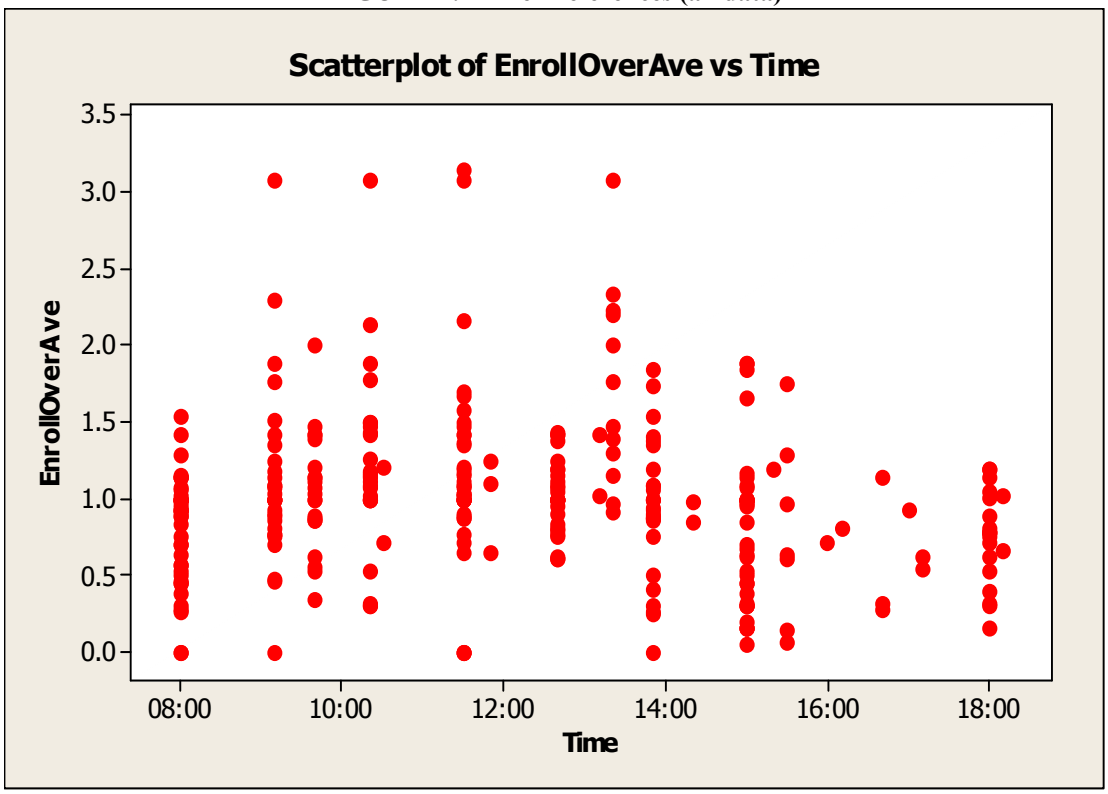
Table 1 shows there are a good array of class times and days represented, as 42.5% of the classes are MWF, 7.5% are evening classes, and good distribution of times (using military clock, coded as date/time variable). The average number of ratings for instructors is fairly high at 7.24, with a standard 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.