

**WORK ETHIC AS A HIGHER-ORDER CONSTRUCT:
EFFECTS OF STUDENT EFFORT ON THEIR
EVALUATIONS OF TEACHING PERFORMANCE**

Kent S. Faight
Ouachita Baptist University
faughtk@obu.edu

ABSTRACT

The topic of work ethic is described here as a higher-order construct. Parts of the higher-order construct are empirically examined within a university context using a mediation approach in order to demonstrate nomological validity. A sample of 1007 end-of-semester student evaluations of teaching (SET) was used. Work ethic was theorized to mediate the effects of factors outside the control of faculty (course and student characteristics) on SET. The effect of grades were also theorized to mediate the work ethic-SET relationship. Results clearly showed the amount of student effort outside of class had consistently positive effects on SET and was not mediated by grades, which was consistent with theory. However, the number of hours spent studying outside of class had no mediating, or direct, effects on SET. A post-hoc analysis found an unanticipated interaction between hours studying and GPA. Poorer students did not like being pushed to study more (which they took out on professors with lower SET), but better students responded positively to being pushed harder. Initial support for the higher-order construct of work ethic was demonstrated.

Introduction

"You can lead a horse to water, but you can't make it drink."

Teaching college students is complex. Teaching is the "leading to water" that professors do. The "drinking" is the learning students do. There are methods that can be used by professors to teach, but this does not mean the students learn. Evidence of the difference between teaching and learning can be observed by noting changes in the accreditation standards of AACSB. So the real difference between teaching and learning is how much the student is involved in their own education. In short, it is how much effort the student puts into it.

Work ethic has been a topic of interest for a long time. So, what is it and how does it work? Here a higher-order construct of work ethic is developed and partially examined within a university context. Rather than examine the extent to which students learn, this study analyzes the authority figure in the education process -- the professor's teaching. So, the issue examined is the extent to which the work ethic of the student affects how they rate their professors.

Literature Review

Max Weber developed the concept of work ethic, specifically the "Protestant Work Ethic," with the publication of his *The Protestant Ethic and the Spirit of Capitalism* in 1905. Weber argued that Protestant countries were more capitalistic because their religion was more open to the idea that making a profit was good than Catholic nations. Over time the religious and western cultural connotations of the concept were minimized.

In his best-seller *The World is Flat* (2005) a century later, Thomas L. Friedman argued one of the reasons for problems with American education relative to the rest of the world was an "Achievement Gap." This was his observation that American students simply did not have as strong a work ethic as non-American students, especially compared to students from BRIC countries (Brazil, Russia, India, and China). Friedman's observation is a more recent observation that parallels previous work (Whyte, Jr., 1956 and Vecchio, 1980). All suggested a decline in the tendency towards hard work in the United States over time.

Studies of college students have shown a number of work-related variables to be related. Adams and Moore (2007) examined a large database containing more than 40,000 students and found that students with a large amount of debt (\$6,000 or more in credit card debt) also had lower GPA, higher BMI, exercised vigorously less, and engaged in more high-risk sexual behavior. Similarly, Kobayashi (2009) found a negative correlation between GPA and BMI (i.e., better students were less fat while weaker students were more fat) among a sample of American and Japanese students. While work ethic is generally expected to relate to better performance, Bottin (1977) found an unanticipated *negative* correlation between work ethic and accounting class grades. This suggests that a variety of measures of college students are related systematically. We believe the construct underlying these connections is work ethic.¹ Our assumption is that work ethic manifests itself in multiple areas of a person's life, not just in at a person's job. We propose a higher-order construct as an integrating conceptualization tying together these various measures, as indicators of work ethic.

A Higher-Order Construct of Work Ethic

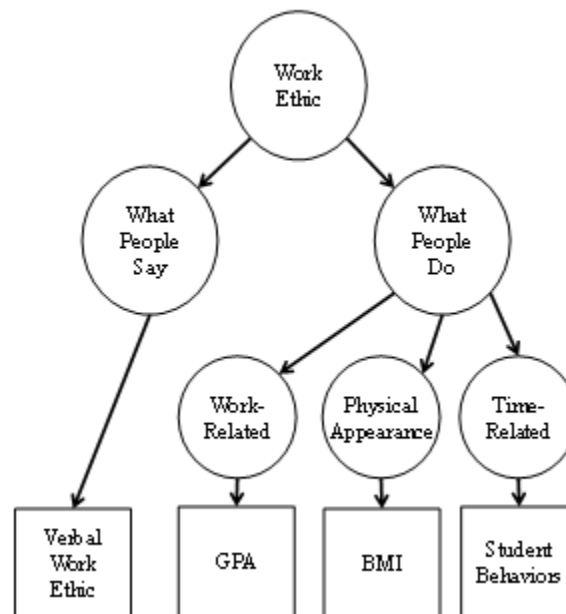
Based on our assumption that work ethic manifests itself in various aspects of one's life, we offer Figure A as the theoretical framework guiding the general approach of this study. Figure A depicts work ethic as a higher-order construct that is one part "What People Say" and one part "What People Do," which are manifested by various indicators. "What People Say" is directly manifested by "Verbal Work Ethic," while "What People Do" is further divided into the sub-dimensions of "Work-Related," "Physical Appearance," and "Time-Related." We believe "Work-Related" hard work is manifested by strong performance at their chosen task. For students, this is indicated by GPA (expected to be higher for hard workers). One indicator of "Physical Appearance" is the degree of "fatness" of one's body, and can be measured by BMI (anticipated to be lower for hard workers). Finally, where one spends their time is believed to indicate the importance of the activity to the person, so hard working students are expected to spend time in more productive ways ("Student Behaviors"). Direct examination of our theory has more fully been conducted in another study (Faught and Graham, 2010), but we discuss it here to more fully place this study in the appropriate context of our thoughts and research agenda.

Assessing the Nomological Validity of "What People Do"

Cronbach and Meehl (1955) introduced the concept of "nomological validity" as one way to demonstrate the validity of a construct. Nomological validity means that constructs will have specific patterns of relationships with antecedents and effects that are unique. If a construct is valid, then its patterns of antecedent and effect relationships will be different than the pattern of relationships of other constructs. When visually depicted, a construct that has nomological validity has a different network of relationships than the network of another construct. **Figure B** is provided with "antecedents" and "effects" noted as a visual reference for this study.

Baron and Kenny (1986) developed a way to assess partial and full mediation which has become the standard methodology in the management literature. First, the relationship between the independent variable (IV) and mediating variable (mediator) is determined (depicted as "a" in **Figure B**). Then the relationship between the IV and dependent variable (DV) is determined (depicted as "b" in **Figure B**). Finally, the relationship between the mediator and DV is determined (depicted as "c" in **Figure B**). While determining this last relationship (c) the IV-DV relationship (b) is of particular importance. If b becomes smaller when c is determined, then this is evidence of partial mediation. If b becomes non-significant when c is determined, then this is evidence of full mediation. **Figure B** shows 2 mediator variables (work ethic and grade equity) as upper and lower portions of the box. The IV's for work ethic are course characteristics and student characteristics (these IV's are also considered antecedents here). The IV's for grade equity are student effort and effort hours. Both mediators, and their theorized IV's/antecedents, are believed to affect the same DV/effect (teaching performance).

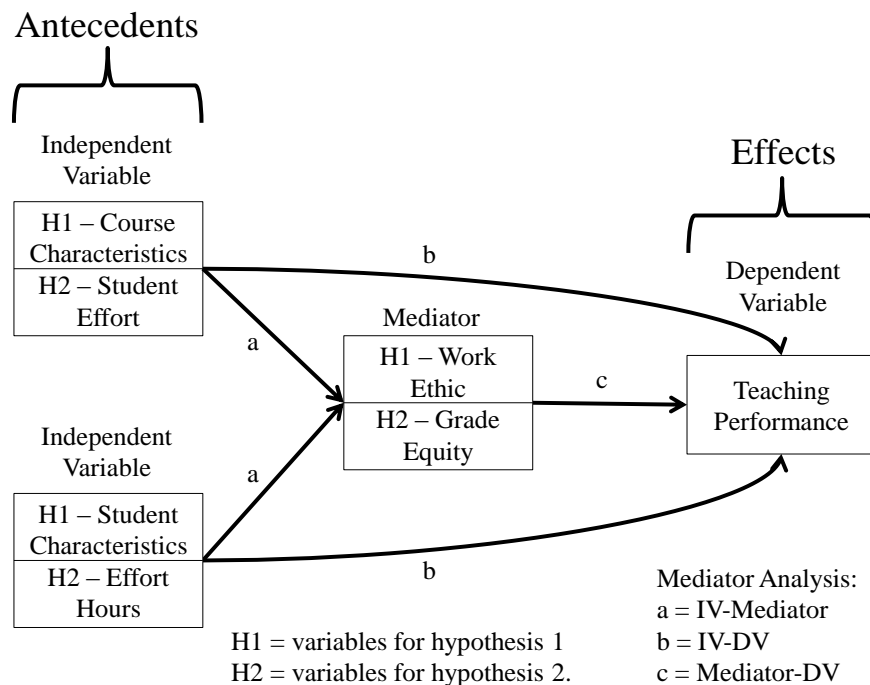
Figure A
Work Ethic: A Higher-Order Construct



One of the most common ways of evaluating the teaching of university classes is having students evaluate their professors. This is frequently referred to as "students evaluating teaching" (SET),

but is labeled here as "Teaching Performance." The logic of these evaluations is straightforward, ask the students what they think because they are in the best position to know what the professor did and how they felt about it. However, evidence had shown these evaluations can be affected by factors unrelated to teaching and sometimes beyond the control of the professor. Not surprisingly, the easiness of the course has been found to be positively correlated with SET (Sonntag, Bassett, and Snyder, 2009). Another study (Seiler, Seiler, and Chiang, 1999) found GPA was positively correlated with SET, while being a required course lowered SET. A result that may or may not have been anticipated, found actual and anticipated grades for a course were unrelated with SET (Moore, 2006). A recent study (Freng and Webber, 2009) even found the non-teaching-related factor of physical attractiveness (i.e., "hotness") to be positively related to SET. While the validity of SET may be questioned by faculty, our measure of it in the context of this study makes sense in that it is a reasonable measure of their perceived experience in the course. If the student worked hard in the course and was fairly rewarded for their efforts, then we would anticipate this would positively affect SET.

Figure B
Mediation Framework: A Nomological Validity Approach



Hypotheses

Based on the discussion in the previous section we offer the following hypotheses:

Hypothesis 1: Work ethic mediates the relationship between course and student characteristics on evaluations of teaching performance.

Hypothesis 1a: Student effort mediates the relationship between course and student characteristics on teaching performance.

Hypothesis 1b: Effort hours mediates the relationship between course and student characteristics on teaching performance.

Hypothesis 2: Grade equity mediates the relationship between work ethic and evaluations of teaching performance.

Hypothesis 2a: Grade equity mediates the relationship between student effort and teaching performance.

Hypothesis 2b: Grade equity mediates the relationship between effort hours and teaching performance.

Method

Sample and Data Sources

The sample consisted of 1007 student evaluations from 15 different professors over 3 semesters at a small, private university in the southern United States. These evaluations were distributed towards the end of the semesters (Fall 2005, Spring 2006, and Spring 2006) to students by the faculty as part of the course evaluation procedure. This particular evaluation was new to the university, having been developed by an internal faculty committee on which the author was a member. The author was intimately involved in the development of the new survey, including initial statistical analysis to assess psychometric properties. The survey was pilot tested on 15 faculty members' evaluations, whose data was used here only after informed consent of each faculty member was obtained.

The full sample of evaluations obtained was 1024. Of these, 17 were from 1 hour classes (i.e., science labs). As a pilot test, some of the science faculty had students complete evaluations during labs. The students were supposed to respond based on teaching in the lab *as well as* the class. Feedback to the internal faculty committee from science faculty questioned whether or not students were evaluating both the lab and the class. Since there was some confusion as to what students might be evaluating, these 17 responses were not included in any analyses here as it would be unrealistic to believe evaluations of a one-hour course would be comparable to evaluations of a three-hour course.

Measures - Dependent Variable

Teaching Performance was measured as the mean of four items concerning the extent to which the student rated the course and instructor (e.g., "Overall, this course is among the best I have ever taken.") as well as the extent to which the student would recommend the course and instructor (e.g., "I would recommend this instructor to a close friend."). Responses on each item were obtained on 5-point scale with Likert-type agree/disagree response options (1 "Strongly Disagree" to 5 "Strongly Agree"). Inter-item reliability was very good ($\alpha = .90$).

Measures - Independent Variables

Two types of independent variables were used: course characteristics and student characteristics. Course characteristics were determined based on the course label while student characteristics were items included on the evaluation completed by the student.

Course Characteristics. *Gen Ed* (short for "general education") was measured as a single item based on the university "department" code for the course. Courses labeled with the code for general education courses were given a value of "1," and all other courses were labeled "0." *Course Level* was measured as a single item based on the university "course number" code for the course. Courses labeled with freshman-level and sophomore-level codes were given a value of "1," while junior- and senior-level courses were labeled "2." *Classes/Week* (abbreviation for "classes per week") was measured as a single item based on the university "days" code for the course. Courses labeled "MWF" were given a value of "3," and all other courses were labeled "2" regardless whether they were "TTH" or "MW." *Morning Class* was measured as a single item based on the university "time" code for the course. Courses listed at times prior to noon were given a value of "1," and courses listed at times from noon onward were given a value of "0."

Student Characteristics. *Classification* was as a single item where students were asked to check a box next to the appropriate answer to the question, "What is your current classification?" Responses were coded 1 (freshman), 2 (sophomore), 3 (junior) and 4 (senior). Responses to an "other" option (N=8), were excluded from the analysis. *GPA* (abbreviation for "grade point average") was a single item where students were asked to check a box next to the appropriate answer to the question, "What is your approximate **overall** GPA?" (emphasis included). Responses were coded 1 (less than 2.00), 2 (2.00-2.49), 3(2.50-2.99), 4 (3.00-3.49), and 5 (3.50 and above). *Class Difficulty* was as a single item where students were asked to check a box next to their answer to the question, "Which of the following best describes this class?" Responses were coded 1 (a very easy class), 2 (an easy class), 3(an average class - neither easy nor hard), 4 (a hard class), and 5 (a very hard class).

Measures - Mediating Variables

Two types of mediating variables were used: work ethic and grade equity. Work ethic was measured both as a frequency of specific class-related behaviors and the number of hours spent studying outside class, while grade equity was a created variable based on two items.

Work Ethic. *Student Effort* was initially measured with 5 items using a frequency 5-point Likert-type scale (1 "Never," 2 "Rarely," 3 "Sometimes," 4 "Often," and 5 "Always") that included a "Not Applicable" (9) option. Not applicable responses were very few (N=8 for the first 4 items below and N=35 for item 5 below), and were removed from the analyses. The 5 items, conceptualized as behaviors hardworking students would report highly on, were the following:

1. "I attended class."
2. "I thoroughly prepared for class."
3. "When I attended class, I was actively engaged (e.g., listened attentively, contributed to class discussions, answered questions, etc.)."
4. "I stayed up-to-date on all course work."
5. "I sought help when I needed it."

Initial inter-item reliability analysis (Cronbach's α) was poor. Upon review, it was clear that the first question ("I attended class") showed virtually no variation (i.e., all students reported "always" attending class), so this item was removed due to evidence of constrained variance.

The fifth question ("I sought help when I needed it") was also found to decrease reliability, and was removed based on the assumption that many hardworking students may not have needed help (N=35 of "not applicable" responses), so they would rate this item low in contrast to items 2-4. The final measure of *student effort* was the mean of 3 items (items 2-4 above). Inter-item reliability was adequate for a first measure of the construct ($\alpha = .69$) according to Nunnally (1967).

Effort Hours was measured as a single, open-ended item ("Not counting class time, how many hours per week did you spend working for your grade?"). A frequency analysis of these responses showed a minimum of 0, maximum of 40, mean of 4.38, median of 3, mode of 2, and standard deviation of 4.02. This author was highly skeptical that a student would work 40 hours (i.e., the equivalent of a full-time job) for a single class, so inspection of the frequency analysis showed that 98.1% of respondents reported working 0-15 hours for the class. Therefore, responses greater than 15 hours were dropped (N=19). Further frequency analysis showed a minimum of 0, maximum of 15, mean of 4.02, median of 3, mode of 2, and standard deviation of 3.04. Neither the median nor mode were changed, while the mean and standard deviation decreased substantially. The revised measure is believed to be far more accurate.

Grade Equity. *Grade Equity* was conceptualized to parallel "pay equity" (Adams, 1965) as a ratio of "rewards" divided by "work." Students answered two items using a scale used to report college grades and calculate GPA (0 "F," 1 "D," 2 "C," 3 "B," and 4 "A"). The two items, conceptualized as anticipated reward and deserved reward respectively (i.e., what one should receive due to one's work contributed), were the following:

1. "What grade do you **expect** to earn in this course?" (emphasis included)
2. "What grade do you **feel** accurately reflects your performance in the class?" (emphasis included)

Grade Equity was created as the simple ratio of item 1 divided by item 2. When both items were identical, the value was 1 and represented *equitable compensation* (i.e., fairly rewarded). When item 1 was larger than item 2, then the value was greater than 1 and represented *over-compensation* (i.e., unfairly rewarded to the benefit of the student). When item 1 was smaller than item 2, then the value was less than 1 and represented *under-compensation* (i.e., unfairly rewarded to the detriment of the student). One interesting and noteworthy observation was that the range of responses to items 1 and 2 did not include any "F's" (0's), which would have made "0" and "indeterminate" ends of the scale. Apparently F's were outside the range of possibility for these students.

Statistical Analysis

In addition to descriptive and zero-order correlation analyses, the primary analyses were hierarchical multiple regressions (Cohen and Cohen, 1983). The Barron and Kenny (1986) methodology was used for the analysis of mediation.

Results

Descriptive statistics on all variables are shown in **Table 1**. The table also shows the zero-order correlations among all measures and inter-item reliability estimates where appropriate.

Table 1
Descriptive Statistics and Correlations

	Mean	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
<i>Dependent Variables</i>													
1. Teaching Performance	3.96	.97	(.90)										
<i>Mediating Variables</i>													
2. Student Effort	4.16	.65	.31**	(.69)									
3. Effort Hours	4.02	3.04	-.04	.19**									
4. Grade Equity (Expected/Deserved)	.99	.22	.15**	-.13**	-.17**								
<i>Independent Variables - Course Characteristics</i>													
5. Gen Ed Course (Y/N)	.31	.46	-.18**	-.21**	-.21**	.03							
6. Course Level (Upper/Lower)	1.38	.49	.10**	.06	.01	-.04	.34**						
7. Classes Per Week (2 or 3)	2.57	.50	.08**	-.01	.12**	.12**	-.03	-.07**					
8. Morning Class (Y/N)	.71	.46	.09**	.05	.02	.07*	-.22**	-.33**	.26**				
<i>Independent Variables - Student Characteristics</i>													
9. Classification	2.46	1.17	.09**	.04	-.05	-.04	-.36**	.63**	-.04	-.12**			
10. GPA	3.99	1.05	.05	.12**	-.01	.02	-.06*	-.04	.09**	.11**	-.15**		
11. Class Difficulty	3.59	.96	-.10**	-.06	.46**	-.21**	-.14**	.07*	-.03	-.08**	-.03	-.07*	

* p<.05; ** p<.01; N's = ranged from 937 to 1007; coefficient alphas are reported in the diagonal where appropriate.

Means and (Standard Deviations)

The mean of teaching performance was 3.96 (0.97) which translated to an "Agree" average varying from "Undecided" to "Strongly Agree." This shows that teaching performance evaluations were skewed upward rather than normally distributed as a mean of 3 on a 5-point scale. The mean of GPA was 3.99 (1.05) which translated to an average of "3.00-3.49" varying from "2.50-2.99" to "3.50 and above." The mean for class difficulty was 3.59 (0.96) which translated to midway between "An average class" and "A hard class" varying from "An easy class" to "A very hard class." The mean of student effort was 4.16 (0.65) which translated to slightly over "Often" varying from "Sometimes" to "Always." The mean of effort hours was 4.02 (3.04) which translates to about 4 hours studying per week varying from a low of 1 hour to a high of 7 hours. The mean of grade equity was 0.99 (0.22) which translated to *equitable compensation*.

Zero-order Correlations of Course and Student Characteristics on Teaching Performance

All of these were statistically significant in the direction teachers would anticipate, with the exception of GPA being non-significant (.05). Gen Ed (-.18) and class difficulty (-.10) were negatively correlated, while upper level courses (.10), morning classes (.09), and classes that met 3 times per week (.09), and classification (.09) and were positively correlated. The pattern of correlations is consistent with the stereotype that required courses (gen ed classes taken by freshman and sophomores) were lower than non-required courses (non-gen ed classes taken by juniors and seniors). The negative correlation also shows harder classes result in lower evaluations while easier classes increase evaluations.

Zero-order Correlations of Course and Student Characteristics on Work Ethic

Several of these were statistically significant in the anticipated direction. Gen Ed was negatively correlated (-.21 for both measures of work ethic), while GPA was positively correlated with student effort (.12). Better students worked harder but students worked less for Gen Ed classes. Gen Ed was also negatively correlated with effort hours, while classes per week was positively correlated (.12). The strongest correlation was between class difficulty and effort hours (.46) indicating that the difficulty of the class was strongly related to the number of hours the student spend studying for it. The pattern of correlations is consistent with the stereotype that required courses (Gen Ed) and harder classes (class difficulty) result in lower evaluations than non-required courses and easier classes.

Zero-order Correlations of Course and Student Characteristics on Grade Equity

Only two factors were significantly related with grade equity. Classes that met three times per week were positively correlated (.12) and class difficulty was strongly correlated (-.21). A positive correlation means classes meeting two times per week were under-compensated while classes meeting three times per week were over-compensated. The strong negative correlation for class difficulty means highly difficult classes were under-compensated while easy classes were over-compensated.

Zero-order Correlations Among Mediating Variables and on Teaching Performance

The correlation between student effort and effort hours (.19) indicates a significant relationship, but far short of a perfect correlation (1.0). Hence, this suggests these are two unique measures of work ethic consistent with our theory (**Figures A & B**). Both student effort and effort hours were negatively correlated with grade equity (-.13 and -.17 respectively), indicating that higher work ethic was under-compensated and lower work ethic was over-compensated. Student effort was strongly correlated with teaching performance (.31), showing that the more effort students invested *outside of the classroom on their own* affected how they evaluated the teacher. This correlation was consistent with our overall theory (**Figure A**) and the specific approach to this study (**Figure B**). Grade equity was also positively correlated (.15), indicating that under-compensated students rated teachers lower and over-compensation increased evaluations. This correlation is consistent with the stereotype that higher grades do increase evaluations. The insignificant correlation between effort hours and teaching performance was *unanticipated*, and

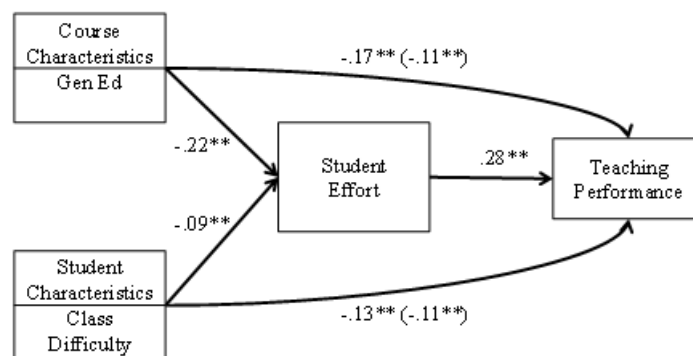
coupled with the strong student effort-teaching performance correlation provides more evidence student effort and effort hours are each unique indicators of work ethic.

Regression Analyses

Prior to hypothesis testing, we ran a regression analysis to determine the effects of all the *Course* and *Student Characteristics* on *Teaching Performance* (b in **Figure B**). Three of these were non-significant (*Morning Class*, *Classification*, and *GPA*), so they were not included in any of the hypothesis testing regression analyses. The remaining characteristics variables (*Gen Ed*, *Course Level*, *Classes/Week*, and *Class Difficulty*) were always entered as the first block of variables in the following hierarchical regression analyses.

The analyses to test *Hypothesis 1* (1a only) are depicted in **Figure C**. Hierarchical multiple regressions (Cohen and Cohen, 1983) were used to analyze the effects of *Course* and *Student Characteristics* on *Student Effort* and on *Teaching Performance* following the Baron and Kenny (1986) mediation methodology depicted in **Figure B**. Results showed a single measure of *Course Characteristics* (*Gen Ed*) and a single measure of *Student Characteristics* (*Class Difficulty*) were negatively related with *Student Effort* (a's of $-.22$, $p < .01$ and $-.09$, $p < .01$ respectively), meaning that students in general education classes or harder classes reported less effort. *Gen Ed* and *Class Difficulty* were also both negatively related with *Teaching Performance* (b's of $-.17$, $p < .01$ and $-.13$, $p < .01$ respectively), meaning that students in general education classes or in harder classes evaluated their teachers lower. *Student Effort* was positively related with *Teaching Performance* ($.28$, $p < .01$) and the b's for *Gen Ed* and *Class Difficulty* on *Teaching Performance* **decreased** ($-.11$, $p < .01$ for both), meaning the more effort the student exerted towards the class the less effect being in a general education or hard class had on their evaluations of the teacher. These analyses provide evidence for a *partial mediating* effect of *Student Effort* that supports *Hypothesis 1a*. Regressions with *Effort Hours* as the mediator showed c to be non-significant providing evidence to reject *Hypothesis 1b*.

Figure C
Testing Hypothesis 1a: A Visual Approach



Mediator Analysis:

a = Gen Ed ($-.22$) & Course Level ($-.09$) both negative; Model $R^2 = .05^{**}$

b = Gen Ed ($-.17$) & Class Difficulty ($-.13$) both negative; Model $R^2 = .06^{**}$

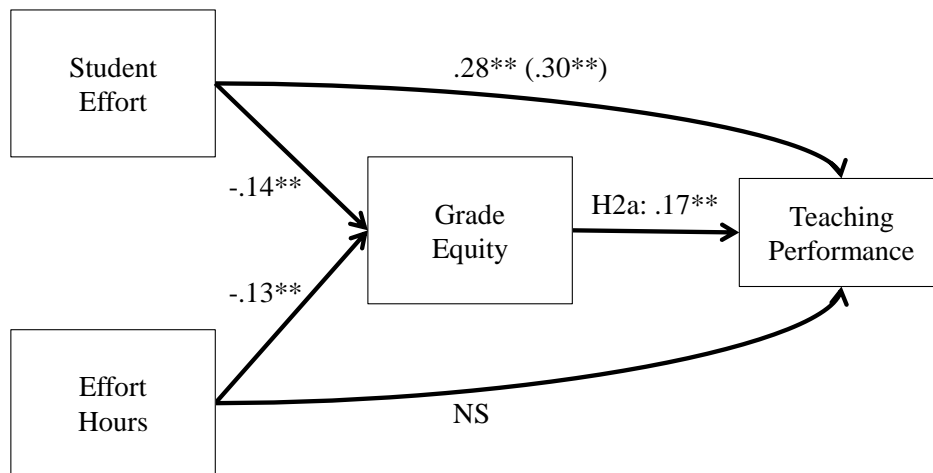
c = Student Effort ($.28$) – changes to b noted in (parentheses)

Gen Ed ($-.11$) & Class Difficulty ($-.11$) **decreased**; Model $R^2 = .13^{**}$

**= $p < .01$

The analyses to test *Hypothesis 2* (2a and 2b) are depicted in **Figure D**. Hierarchical multiple regressions (Cohen and Cohen, 1983) were used to analyze the effects of *Student Effort* and *Effort Hours* on *Grade Equity* and on *Teaching Performance* following the Baron and Kenny (1986) mediation methodology depicted in **Figure B**. Results showed both *Student Effort* and *Effort Hours* were negatively related with *Grade Equity* (a's of $-.14$, $p < .01$ and $-.13$, $p < .01$ respectively), meaning that students that exerted more effort or hours to a class felt *under-compensated*. *Student Effort* was positively related with *Teaching Performance* (b of $.28$, $p < .01$) but *Effort Hours* was unrelated to *Teaching Performance*, meaning that students that exerted more effort towards a class (but not in terms of hours) evaluated their teachers higher. *Grade Equity* was positively related with *Teaching Performance* ($.17$, $p < .01$), meaning the more students were *over-compensated*, the higher they evaluated the teacher. For this last regression, the b for *Student Effort* on *Teaching Performance* **increased** ($.30$, $p < .01$ for both), meaning the perceptions of grade equity had no mediating effect on student's evaluations of the teacher. These analyses provide no evidence for a mediating effect of *Grade Equity* that supports *Hypothesis 2a* or *2b*.

Figure D
Testing Hypothesis 2: A Visual Approach



Mediator Analysis: Hypothesis 2a
 a = Student Effort ($-.14$) negative; Model $R^2 = .08^{**}$
 b = Student Effort ($.28$) positive; Model $R^2 = .13^{**}$
 c = Grade Equity ($.17$) – changes to b noted in (parentheses)
 Student Effort ($.30$) **increased**; Model $R^2 = .15^{**}$

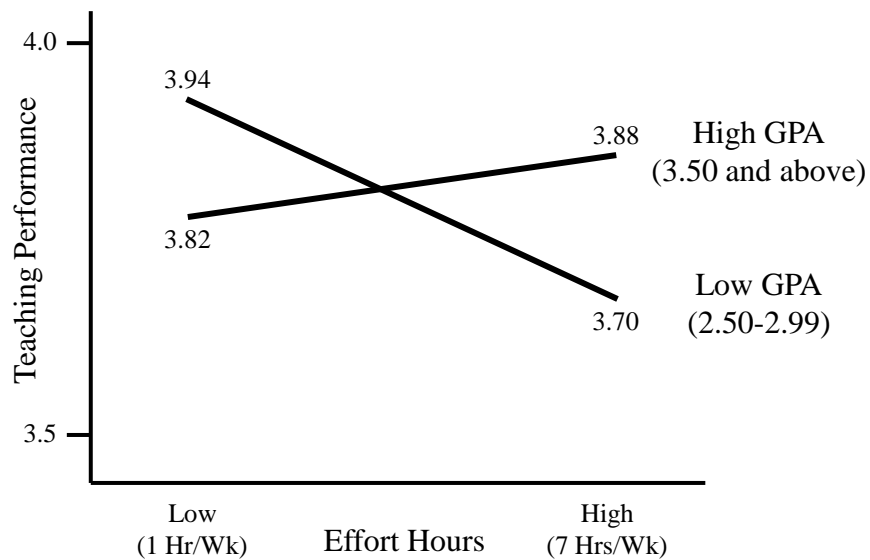
**= $p < .01$; NS = non-significant

Post-Hoc Analysis of Effort Hours

The insignificant effects of effort hours and GPA on teaching performance were inconsistent with our theory (**Figure B**), to which we had no a priori explanation. As a purely exploratory analysis, we conducted a follow-up hierarchical regression analysis (Cohen and Cohen, 1983). *Course* and *Student Characteristics* were included along with *GPA* and *Effort Hours*. We created

an interaction term for *GPA* and *Effort Hours* and entered it as the last block, which was significant. To visualize the interaction term, it was plotted by calculating teaching performance from the regression results. Values inserted into the equation were 1 standard deviation above and below the means and are depicted in **Figure E** (1,1 for high *GPA* and high *Effort Hours*; 1,-1 for high *GPA* and low *Effort Hours*; -1,1 for low *GPA* and high *Effort Hours*; and -1,-1 for low *GPA* and low *Effort Hours*). **Figure E** shows that low *GPA* students (those with GPA's "2.50-2.99") rated teaching performance highest when they only had to study 1 hour a week for the class, and rated teaching performance much lower when they had to study seven hours a week. High *GPA* students (those with GPA's "3.50 and above") rated teaching performance lower when they studied only 1 hour for the class than they did when they studied seven hours a week. This *unanticipated* result showed pushing weaker students to work harder led to lower evaluations, while the same push on stronger students increased evaluations.

Figure E
Examining the Effort Hours X GPA Interaction: A Visual Approach



Post-Hoc Analysis: Plotting the Interaction
 High values set at 1 STD above mean; Low values set at 1 STD below mean.

Discussion

The Big Picture - Nomological Validity of our Higher-Order Construct of Work Ethic

Our higher-order construct depicted in **Figure A** argues there are different dimensions and manifestations of work ethic. In order to begin substantiating this claim we applied a mediation framework to demonstrate nomological validity. We theorized work ethic would mediate the effects of course and student characteristics on SET. We also theorized grades would mediate the work ethic-SET relationship. Regression analyses clearly showed *Student Effort* (defined as work prior to, during, and after class) did mediate as expected, providing support for *Hypothesis 1a*, but that grades did not affect student effort's effects on SET (*Hypothesis 2a* rejected). However,

when work ethic was measured as the number of hours spent studying for the class (*Effort Hours*) work ethic did not have effects on SET as anticipated (*Hypotheses 1b* and *2b* rejected). Our post-hoc interaction analysis showed *Effort Hours* interacted with *GPA* such that better students rewarded faculty with higher SET when pushed to study more, but weaker students punished faculty with lower SET when pushed to study more.

Although *Student Effort* and *Effort Hours* were significantly correlated, the pattern of antecedents and effects for *Student Effort* differs from the pattern of antecedents and effects for *Effort Hours*. These results, therefore, demonstrate the nomological validity of these two work ethic measures being unique. This provides initial support for our higher-order construct of work ethic being manifested by unique indicators of different dimensions (*Figure A*). We hope further research will examine other indicators and dimensions to confirm or reject our construct.

The Narrow Picture - Assessing the Effects of Student Work Ethic in a University Context

The pattern of results in this study clearly shows SET are affected by factors outside faculty control. *Course Characteristics*, specifically being a general education course (*Gen Ed*), showed consistently negative effects on SET. *Student Characteristics*, specifically being perceived as a harder class (*Class Difficulty*), also was found to have consistently negative effects on SET. As theorized, the work done by the student as part of their learning process (*Student Effort*) was consistently positively related with SET and this effect was direct rather than mediated by grades (*Hypothesis 2a* rejected). This means that the work ethic of the student, which is under the control of the student rather than the professor, significantly affects how the student evaluates the professor. The interesting interaction between the type of student in the class (*GPA*) and the work ethic measure of *Effort Hours* (*Figure E*) shows that these two factors, both of which are beyond the control of the faculty, significantly affect the professor's SET.

Evaluations of college professor's teaching (SET) is a common university practice. The assumption is that good evaluations reflect good teaching, and bad evaluations reflect bad teaching. The results provided here demonstrate the fallacy of this assumption. Factors beyond faculty control should be considered also, as these analyses clearly demonstrate the need for. In order to have a more complete understanding of the factors that affect SET, it will be necessary to take into account factors that are and are not under the control of faculty. We hope we, and others, will be able to conduct these types of analyses in the future.

Level-of-Analysis Limitation

There is an important limitation that must be addressed here regarding any actions that might be taken based on these analyses. While the sample size for this study was fairly large (N=1007), each of these responses were obtained from students. It is reasonable to take these findings to understand how students think, but it is unreasonable and wrong to take actions towards professors based on these data. These data were obtained from a small sample of faculty (N=15). Prior to taking action towards faculty, it is imperative to understand the statistical relationships at the *level-of-the-class* and/or the *level-of-the-professor*. Simply because a statistical relationship exists for students does not mean this statistical relationship is *isomorphic* (i.e., the relationship is the same) at higher levels of analysis (for a thorough discussion of level-of-analysis issues see

Klein and Kozlowski, 2000). Faculty teach multiple students in their classes. Simply because of numbers, larger classes can exert an influence on the data that may be misleading. To more accurately understand statistical relationships at higher levels (for the class or the professor), multiple student responses must be *aggregated* (e.g., averaged) to whatever higher level is of interest. For example, aggregating student responses to the *level-of-the-class* would allow accurate interpretations to determine what classes tend to be most highly rated in terms of SET. This sample is far too small to be able to aggregate and maintain sufficient statistical power for accurate interpretations. We point out this limitation because too often analyses of student data are inappropriately used to make policy recommendations for faculty (e.g., Seiler, Seiler & Chiang, 1999).

References

- Adams, J. S. 1965. "Inequity in social exchange." *Advances in Experimental Social Psychology* 62: 335-343.
- Baron, R. M. & D. A. Kenny. 1986. "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of Personality and Social Psychology* 51 (6): 1173-1182.
- Cohen, J. & Cohen, P. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Second Edition. Hillsdale, New Jersey. Lawrence Erlbaum Associates, Publishers.
- Faught, K. S. & D. Graham. 2010. "Work ethic as a higher-order construct: Dean Wormer was right - lazy, fat, and stupid go together (but so do hard-working, fit, and smart)." (In preparation).
- Freng, S. & D. Webber. 2009. "Turning up the heat on online teaching evaluations: Does "hotness" matter?" *Teaching of Psychology* 36 (3): 189-193.
- Cronbach, L. J. & P. E. Meehl. 1955. "Construct validity in psychological tests." *Psychological Bulletin* 52 (4): 281-302.
- Klein, K. J. & S. W. Kozlowski (Eds.). 2000. *Multilevel Theory, Research and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco. Jossey-Bass.
- Moore, T. 2006. "Teacher evaluations and grades: Additional evidence." *Journal of American Academy of Business* 9 (2): 58-62.
- Nunnally, J. C. 1967. *Psychometric Theory*. USA. McGraw-Hill
- Seiler, M. J., Seiler, V. L. & D. Chiang. 1999. "Professor, student, and course attributes that contribute to successful teaching evaluations." *Financial Practice & Education* 9 (2): 91-99.

Sonntag, M. E., Bassett, J. F. & T. Snyder. 2009. "An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com." *Assessment & Evaluation in Higher Education* 34 (5): 499-504.

Endnote

¹ I use the term "we" throughout this paper because a former OBU student, Daniel Graham, began this project as a Directed Study with me. Daniel was instrumental in helping spur the project forward. We collectively arrived at the assumption that work ethic manifests in multiple areas of a person's life, which is the basis on which our higher-order construct (Figure A) was developed. I am grateful for all of Daniel's hard work. He would be one of those that scored high on work ethic (*Student Effort*).