Title: College Academic Performance and External Assessment: Has the Relationship Changed?

Dr. Frederick Tannery
Associate Professor of Economics
School of Business
Slippery Rock University

Dr. Sunita Mondal
Assistant Professor of Economics
School of Business
Slippery Rock University

Growing political pressure from state governments along with strong endorsement of accrediting organizations has led many colleges and universities to adopt formal assessment programs. These programs include internal and external instruments; former including measures generated by the institution, latter a test administered by an outside organization. This research compares internal and external assessments at one institution and how they have changed over time. We find reasonable close relationship between internal and external assessments in the early years of this period became weaker in the later years. Widening gap between the internal and external assessment outcomes comes is due to a growing difference in how student characteristics affect each assessment outcome.

INTRODUCTION

Higher education is in a unique position that there is no systematic effort to evaluate student learning. While evaluating learning is confounded by the differences in majors and curriculum, there is also institutional resistance to structured evaluation efforts. In their defense, selective and usually expensive institutions along with public research institutions argue that they pass a market test as they maintain enrollment, charge tuition that is increasing in real terms and attract students from the upper end of the ability distribution measured by scores on standardized tests taken during high school. Colleges and universities offering enrollment to students much further down the ability distribution contend that they provide a high value added that enables high school graduates to move into the economic mainstream, and should not be subject to the same standard as selective institutions.

Despite these objections, there is growing political pressure from federal and state governments along with the strong endorsement by accrediting agencies that have led many institutions to adopt formal assessment of student learning. Assessment efforts are decentralized as evaluating student learning is left to academic departments to establish the criteria of what students learned in their field of study. Strategies frequently have internal and external measures of student performance. The first includes college grades, student surveys of course effectiveness, and alumni surveys noting post-college employment and satisfaction with their college experience. External measures are created outside the institution and test knowledge within a major. The Educational Testing Service (ETS) has developed a family of tests designed to measure student mastery of specific disciplines. Our interest here is in the Major Field Achievement (MFT) in business that tests knowledge of the business curriculum.

The purpose of this paper is to compare one internally generated measure - grades with the external measure –test scores and examine how the relationship between them has changed over time. More specifically, we compare overall grade-point average (GPA), and grades in a common business core curriculum (CORE) to MFT scores for students at one medium sized public institution that has administered the test over a 20-year period, although there is a three-year gap in test scores owing to administrative issues and budgetary considerations. We exploit the long study period to show how the grade-test score relationship has changed over time. To our knowledge, no other studies have considered such a long period of time to explore the relationship between the grades and test scores. We first establish the grade-test score relationship in the early part of the period for test takers from 1995 through 2001 by first using GPA and then both GPA and CORE to predict test scores. We then augment grades with other student information related to college experience, pre-college ability and demographic characteristics. We then repeat this analysis for students from the later period and compare the results. Differences over time are due to changes in student characteristics and changes in the influence of the characteristics on outcomes. We use the Blinder-Oaxaca decomposition to isolate each effect along with an interaction term. We provide visual evidence of the decomposition of the role of characteristics and the influence of characteristics on outcomes with kernel densities that highlight the counterfactual distribution of the outcomes based on characteristics in the later period and regression estimates form the earlier period.

We do not suggest that these results can be extrapolated elsewhere, but only document the relationship at one institution and how they have changed over time. As improving grades and falling test scores is part of a national pattern1, these results could be informative about the grade-test score relationship at institutions that enroll students of similar ability.

RELATED LITERATURE

Research reconciling MFT scores with college grades are all based on private information that is not available to the research community. While there are some overall trends, the reader should bear in mind the institution for which the test scores are reported. Unlike other studies that use proprietary college data in discussing topic such as grade inflation, e.g. Johnson (2003), Sabot and Wakemann-Linn (1991), colleges and universities using the MFT are generally not selective private institutions or public research universities.

There is a sizable literature on the MFT in Business that reflects the large number of institutions that administer the test2. Several studies find that students with higher grades do better on the MFT (Allen and Bycio 1997; Terry et al., 2008). Mirchandani, Lynch, and Hamilton (2001) find that scores are higher for men and increase with SAT score, the GPA of transfer students and grades in quantitative courses. They conclude that the SAT score is a dominant variable explaining most of the variation in MFT exam scores. Bycio and Allen (2007) found that SAT scores account for 53 percent of the variation in exam scores, and adding the business core GPA accounts for 65% of the variation in MFT scores. They also find that highly motivated students do better on MFT test3. Bagamery, Lasik, and Nixon (2005) report gender, SAT scores, and grades to be significant determinants of the MFT exam, while they find age, transfer status, and major do not significantly affect the exam outcome. In particular, their study indicates that male students score 8.5 points higher on the MFT, and higher GPA leads to higher MFT scores. Contreras et al. (2010) finds that MFT performance varies among business majors, as accounting and management majors underperforming other majors. Unlike these studies, we not only estimate the determinants of test scores, but also report how the test score-student characteristic relationship has changed over our 20-year study period. The rest of the paper is organized as follows: Section III describes the data set and descriptive statistics. Section IV explains the empirical results for the determinants of performance on the MFT exam. Section V provides a summary and discussion of the results.

DATA

The data for this paper comes from test results for students at one medium-sized state institution who took the Major Field Achievement Test in Business between 1995 and 2014. The test was given to business majors enrolled in the capstone business course, which was exclusively taken in their final year of study. Test scores were merged with information from student records. These records contain detailed information on student activity such as major, credits earned, and grade-point average. It also contains demographic information including age, race, gender and county of residence. Students applying as freshmen report their scores on the SAT, their high school standing and their high school class size. Applicants seeking to enter college as transfer students only report the number of credits that transferred, but neither SAT scores or high school performance. This research includes only freshman applicant who reported SAT scores and high school rank.

TABLE 1
OVERALL DESCRIPTIVE STATISTICS

Variable	Group 1 (1995-2001) Mean	Group 2 (2002-2014) Mean
Test Score	155.96	150.73
GPA	2.92	3.06
CoreGPA	2.79	2.99
Age	18.36	18.49
Female (%)	48.91	42.35
Local (%)	26.81	26.14
Pre-college Ability		
SAT Math	483.95	489.79
SAT Verbal	419.82	408.14
High School Rank	0.31	0.37
Majors (%)		
Accounting	28.62	19.35
Economics & Finance	16.30	17.39
Marketing	25.36	31.76
Management	23.91	30.46
International Business	5.80	1.05
Observations	276	765

While the data spans 20 years, we have no data from 2002 through 2004 due to budgetary considerations and other administrative problems. Thus we break the data into two groups. Group 1 contains those who took the test before the break, from 1995 through 2001, and Group 2 who took the test from 2005 to the present. Table 1 reports the means of selected variables for the two groups. The later period contains many more observations not only because it is three years longer, but mainly due to the test being administered to students in both the fall and spring semesters in the capstone class while it was given in only one semester in the earlier period. The MFT scores is now an important part of the business program assessment. Before 2002, there were no formal assessment efforts.

Among the most notable differences between Group 1 and Group 2 is the drop in the MFT score by about 5.2 points, which is 47% of a standard deviation. At the same time, however, both the overall grades (GPA) and grades in the business core (CORE), which is a common set of courses required on all students increased by 0.14 and 0.20 points respectively. The two groups are similar in age, but Group 1 contains more women. There has also been some shift of students between majors as marketing and management grew by 2.76% and 6.55% respectively, while accounting majors fell by 9.27%.

An easy comparison of SAT scores is compromised by the re-centering of the test scores beginning for tests taken in April of 1995. This lowers the scores for those who entered before 1996 and could also lower scores for those admitted in 1996 as many students take the SAT in their junior year of high school. We see some evidence of the re-centered scores as the math and verbal scores for students admitted in 1997 were 33 and 77 points higher respectively, then math and verbal scores of students entering college in 1995[4]. Some of this difference could be due to improved student quality as the 1997 cohort were at the 29th percentile of their high school class compare to the 35th percentile for the 1995 cohort. While there is always some variation in test scores within the groups, the wide difference between groups suggests that re-centering of scores is the main reason. We adjusted to the re-centering

by deducting 24 points from the math score and 76 points from the verbal score for the cohorts entering in 1996 or later[5].

TABLE 2
DESCRIPTIVE STATISTICS BY MAJOR (1995-2001)

Variable	(1)	(2)	(3)	(4)	(5)
GPA	3.28 (0.41)	3.12 (0.44)	2.98 (0.47)	2.97 (0.45)	3.25 (0.51)
Core GPA	2.61 (0.41)	3.12 (0.61)	2.86 (0.59)	2.89 (0.59)	2.97 (0.64)
Test Score	153.05 (11.73)	154.50 (12.09)	149.80 (9.94)	148.06 (11.16)	152.38 (12.49)
Pre-college ability					
SAT-Math	513.16 (65.49)	505.02 (70.29)	474.60 (66.84)	481.49 (70.29)	492.25 (54.23)
SAT-Verbal	404.34 (56.55)	419.04 (60.23)	403.09 (58.17)	409.02 (67.21)	420.25 (91.33)
High School Rank	0.30 (0.19)	0.36 (0.21)	0.41 (0.21)	0.39 (0.19)	0.38 (0.12)
Number of Observations	148	133	243	231	8

(1) Accounting/Certificate in Accounting
(2) Economics/Finance
(3) Marketing
(4) Management/Human Resource Management/Sustainable Management
(5) International Business

Table 2 compares student performance and pre-college ability by major for the two time periods. Accounting students did better in nearly every measure in early period and enjoyed a better than one-half point (one half of a letter grade) in the CORE compared to marketing students and about nearly 0.6 points higher than management majors. Accounting majors scored 48 points higher on the math part of the SAT than management majors and 38 points higher than marketing majors, and also were at the top one-quarter of their high school class compared to the 38rd and 35st percentile for marketing and management respectively. Finance and economics majors had math SAT scores that were about 23 points lower than those of accounting students, and had CORE grades that were 0.47 points lower than those of accounting students. Finance and economics students did better in high school than marketing or management students but not as well as accounting majors.

Averages by major in Table 3 for Group 2 differ in important ways. Accounting students have the lowest CORE grade, which is better than one-half letter grade below finance and economics students and one-quarter of a letter grade below the others. While accounting students have higher math SAT score than the others and did better in high school, the gap between the other majors' narrows considerably. Comparisons of the two tables suggests that grade inflation is present as grades, with the exception of CORE grades for accounting students increased, while both SAT scores, high school rank and test scores fell.

TABLE 3
DESCRIPTIVE STATISTICS BY MAJOR (2004-2014)

Variable	(1)	(2)	(3)	(4)	(5)
GPA	3.28 (0.41)	3.12 (0.44)	2.98 (0.47)	2.97 (0.45)	3.25 (0.51)
Core GPA	2.61 (0.41)	3.12 (0.61)	2.86 (0.59)	2.89 (0.59)	2.97 (0.64)
Test Score	153.05 (11.73)	154.50 (12.09)	149.80 (9.94)	148.06 (11.16)	152.38 (12.49)
Pre-college ability					
SAT-Math	513.16 (65.49)	505.02 (70.29)	474.60 (66.84)	481.49 (70.29)	492.25 (54.23)
SAT-Verbal	404.34 (56.55)	419.04 (60.23)	403.09 (58.17)	409.02 (67.21)	420.25 (91.33)
High School Rank	0.30 (0.19)	0.36 (0.21)	0.41 (0.21)	0.39 (0.19)	0.38 (0.12)
Number of Observations	148	133	243	231	8

(1) Accounting/Certificate in Accounting
(2) Economics/Finance
(3) Marketing
(4) Management/Human Resource Management/Sustainable Management
(5) International Business

RESULTS

Regression Estimates

The comparison of internal and external assessments is based on the amount of the variation in test scores that can be explained by grades. We first use only GPA as the determinant of test scores before adding CORE grades to the model. We then add other student characteristics including college treatment, demographic characteristics and pre-college ability such as SAT scores and high school rank[6]. Specifically, the model takes the following functional form for the estimation of the larger model:

$$MFT_i = \alpha + \beta_1 GPA_i + \beta_2 CORE_i + \beta_3 X_i + \varepsilon_i \qquad (1)$$

where MFT_i is student i's test score, GPA and CORE have been previously defined, X_i includes the individual-specific covariates mentioned above, α and βs are the parameters to be estimated, and ε_i is the normally distributed error term with a zero mean and a constant variance.

TABLE 4
REGRESSION ESTIMATES OF TEST SCORES (1995-2001)

Variable	(1)	(2)	(3)
Grade-Point Average (GPA)	13.49*** (1.32)	4.31 (2.99)	6.56*** (2.53)
Core Grade-Point Average (CORE)		7.67*** (2.26)	2.75 (1.95)
Fem (1 if female)			-3.62*** (0.97)
Age			-9.49*** (3.06)
Local (1 if from a local county)			0.61 (1.06)
Credits			0.11** (0.06)
SAT Math			0.03*** (0.01)
SAT Verbal			0.06*** (0.01)
HS Rank			2.47 (2.93)
Econfin (1 if finance/economics)			4.13*** (1.5847)
Mktg (1 if marketing)			-0.93 (1.35)
Mgmt (1 if management)			-2.47** (1.37)
Constant	116.49	121.96	89.22
N	276	276	275
Adjusted R^2	0.2732	0.3001	0.5551

* Significant at the 10% level
** Significant at the 5% level
***Significant at the 1% level

Table 4 reports the results for Group 1 students who took the test between 1995 and 2001. The simple regression in column 1 shows a strong relationship between grades and test scores as a one-quarter letter grade improvement in grades leads to a 3.32-point increase in MFT scores. The elasticity at the mean is 0.25, which was also the found in a log specification that is not reported. The single variable also explains 27% of the variance in scores. Adding the core grade point average (CORE) in column 2, not only has a strong positive and precisely estimated impact on the MFT score, but the explanatory power of the model increases to 30%. Furthermore, the GPA is no longer statistically significant. Hence, conditioning on core grades, within major improvement in grades has no significant impact on the test score. While 30% of the variation in MFT scores leaves a large proportion unexplained, this parsimonious model explains over twice the variation in MFT scores than high school rank explains of either SAT score for the same cohort.

The larger model listed in column 3 continues to find a positive impact of grades, but only the effect of GPA reaches conventional level of statistical significance. Conditional on college grades, students who did better on the math and verbal SAT did better on the MFT with both coefficients significant at the 1% level. While high school rank affects college grades, once we control for college grades, high school performance has no significant impact on MFT test scores. Test scores also increase with the number of college credits and vary with the field of study as economics and finance students outperform while management and international business underperform the normalized group - accounting students. Men do better than women and students who were at least 20 year-olds at matriculation did better than their younger counterparts, which would be consistent with the added maturity of students that are not fresh out of high school. The additional variables also switched the significance of the two grades, as GPA is significant but CORE is not. The added variables account for much more of the variation in MFT scores as the adjusted R^2 rises to 0.555, which represents an 83% increase compared to column 2.

TABLE 5
REGRESSION ESTIMATES OF TEST SCORES (2004-2014)

Variable	(1)	(2)	(3)
Grade-Point Average (GPA)	7.36*** (0.85)	3.89*** (1.11)	2.68*** (1.22)
Core Grade-Point Average (CORE)		4.01*** (0.84)	3.48*** (0.82)
Fem (1 if female)			-2.36*** (0.81)
Age			-4.02** (2.18)
Local (1 if from a local county)			0.29 (0.85)
Credits			0.09*** (0.04)
SAT Math			0.02*** (0.01)
SAT Verbal			0.04*** (0.01)
HS Rank			3.38 (2.26)
Econfin (1 if finance/economics)			0.62 (1.24)

Mktg (1 if marketing)			-0.79 (1.15)
Mgmt (1 if management)			-3.14*** (1.16)
IB (1 if international business)			-4.29 (4.08)
Constant	128.18	126.82	98.76
N	765	765	678
Adjusted R^2	0.0885	0.1141	0.2668

* Significant at the 10% level
** Significant at the 5% level
***Significant at the 1% level

Table 5 presents the regression estimates for same model as Table 4, except that it is fit on Group 2 test-takers who took the MFT between 2005 and 2014. The column 1 estimate is still positive and significant although it falls by 45% relative to Group 1 and has an elasticity of .15 compared to .25 for Group 1. The adjusted R^2 is also only .0885, which is much lower than for Group 1. Column 2 adds CORE and both are positive and precisely estimated and together explain 11.41% of the variation in test scores. Both grades are positive and significant in the larger model that includes student characteristics reported in column 3. The other results are similar to the column 3 results in Table 4 for Group 1. Students who did better on the SAT did better on the MFT, men score higher than women and older students score higher than their younger counterparts. For Group 2, only management students underperform accounting students. The most striking difference between the tables is the smaller coefficients and the much lower explanatory power in Table 5. The larger model in column 3 explains only 26.6% of the variation in MFT scores or 48% as much as the model explained for Group 1. The proportional difference in the explained variation between the models using only grades is even larger between the periods. While one can argue that 55% of the variation in MFT scores accounted for by the larger model in the early period shows general agreement between the internal and external assessment measures, we find a much weaker relationship between the two in the later period.

Additional comparisons of internal and external assessment outcomes use the same regression model to explain test scores and grades. Strong similarities in regression estimates that explain tests and grades suggest a close relationship between both measures.

TABLE 6
REGRESSION ESTIMATES OF PERFORMANCE DEPENDENT VARIABLES (1995-2001)

Independent Variables	Test Score	GPA	CORE
Fem	-3.43*** (1.05)	0.05 (0.05)	-0.04 (0.06)
Age	9.61*** (3.36)	0.02 (0.15)	0.09 (0.19)
Local (1 if from a local county)	0.87 (1.17)	0.02 (0.05)	0.05 (0.07)
Credits	0.13*** (0.06)	0.001 (0.003)	0.003 (0.003)
SAT Math	0.03*** (0.01)	0.0003 (0.0003)	0.0009*** (0.0004)
SAT Verbal	0.07*** (0.01)	0.0011*** (0.0003)	0.0010*** (0.0005)
HS Rank	-5.51** (3.02)	-0.78*** (0.13)	-1.04*** (0.17)
Econfin (1 if finance/economics)	1.72 (1.57)	-0.22*** (0.07)	-0.34*** (0.09)
Mktg (1 if marketing)	-2.93*** (1.44)	-0.17*** (0.06)	-0.32*** (0.08)
Mgmt (1 if management)	-5.37*** (1.45)	-0.27*** (0.06)	-0.41*** (0.08)
Constant	111.309	2.53	2.00
N	275	275	275
Adjusted R^2	0.4596	0.3277	0.3579

* Significant at the 10% level
** Significant at the 5% level
***Significant at the 1% level

TABLE 7
REGRESSION ESTIMATES OF PERFORMANCE DEPENDENT VARIABLES (2002-2014)

Independent Variables	Test Score	GPA	CORE
Fem	-1.87^{***} (0.83)	0.08^{***} (0.03)	0.08^{**} (0.04)
Age	3.33 (2.45)	0.02 (0.08)	0.18 (0.12)
Local (1 if from a local county)	0.36 (0.88)	-0.01 (0.03)	0.03 (0.05)
Credits	0.10^{***} (0.04)	0.001 (0.001)	0.004^{***} (0.002)
SAT Math	0.03^{***} (0.01)	0.0009^{***} (0.0002)	0.0010^{***} (0.0003)
SAT Verbal	0.05^{***} (0.01)	0.0009^{***} (0.0002)	0.0005 (0.0003)
HS Rank	-1.83 (2.10)	-0.95^{***} (0.08)	-0.95^{***} (0.08)
Econfin (1 if finance/economics)	0.24 (1.28)	-0.09^{***} (0.05)	-0.04 (0.07)
Mktg (1 if marketing)	-2.19^{**} (1.17)	-0.18^{***} (0.04)	-0.26^{***} (0.06)
Mgmt (1 if management)	-4.59^{***} (1.17)	-0.21^{***} (0.04)	-0.25^{***} (0.06)
Constant	111.31	2.53	2.00
N	678	678	678
Adjusted R^2	0.2051	0.3811	0.1968

* Significant at the 10% level
** Significant at the 5% level
***Significant at the 1% level

Tables 6 and 7 present regression estimates with different measures of performance for students admitted as freshmen for the earlier and later period, respectively. The regression model listed in Table 7 explains less of the variation in MFT scores and CORE grades for Group 2, but more of the variation in GPA. Men and older students do better on the MFT but not on either grade, while student who did better on the SAT had both better test scores and higher grades, the impact on the MFT score is more precisely estimated. The influence of high school rank is particularly notable as, conditional on other regressors, students who did better in high school have higher grades and do better on the MFT with very large and significant impact on grades. Accounting students have higher grades than any other major and did better on the MFT with the exception of economics and finance majors. For Group 2 in Table 7, women have lower test scores but higher grades; students with higher SAT scores did better on all outcome measures and only the estimate of the verbal SAT on CORE grades failing to reach conventional levels of statistical significance. Students with better high school grades did much better on both grade outcomes, but did no better on the MFT. Conditional on SAT scores, characteristics that increase high school grades are still operative when they attend college.

The statistical model accounts for much less of the variation in MFT scores and CORE grades for Group 2 since the explanatory power of the model with test scores as the dependent variable fell from 46% for Group 1 to 20.5% for Group 2 and with CORE grades as the dependent variable the adjusted R^2

was 35.8% and 19.7% for Group 1 and 2 respectively. Somewhat surprisingly, the regression model explains more of the variation in GPA for Group 2 than for Group 1. In particular, compared to Group 1 students, high school rank and math SAT scores have a very large impact on college grades. We have no explanation for these differences. Differences in the explanatory power of the statistical model along with differences in the sign and significance of the regression coefficients presents a clear picture that the data generating process governing MFT scores is different that the on determining grades and that these differences were present during the early years of our study period.

Change Over **Time**

Test scores and grades moved in opposite directions over the study period. Explanations for changes in either the internal or external assessment indicators follows along the lines of standard explanations for differences between groups that is popular in the analysis of wage differences by gender or race that was initially developed by Blinder (1973) and Oaxaca (1973). Rather than ask what women would earn if they had the same characteristics as men, we ask what would be the test scores for Group 2 students if they had the same characteristics as Group 1 students. More formally, we are interested in

$$Y_{1i} = X_{1i}\beta_1 + \varepsilon_{1i} \tag{2}$$

and

$$Y_{2i} = X_{2i}\beta_2 + \varepsilon_{2i} \tag{3}$$

where Y_1 and Y_2 are the outcomes of interest for Groups 1 and 2, X_1 and X_2 represent the characteristics of the subscripted groups, β_1 and β_2 are coefficients, and ε_1 and ε_2 are the error terms of each group. Our interest is in the average growth rate over time which we define as:

$$\begin{aligned} D = \overline{Y}_2 - \overline{Y}_1 &= X_2\beta_2 - X_1\beta_1 \\ &= (X_2 - X_1)\beta_1 - X_1(\beta_2 - \beta_1) + (X_2 - X_1)(\beta_2 - \beta_1) \end{aligned} \tag{4}$$

In equation (4) the first term represents the effect of changing characteristics holding the coefficients constant (the endowment effect), the second term holds the initial characteristics constant and changes the coefficients (the coefficient effect), and the last term shows the simultaneous change in both (the interaction effect).

TABLE 8
OAXACA DECOMPOSITION RESULTS

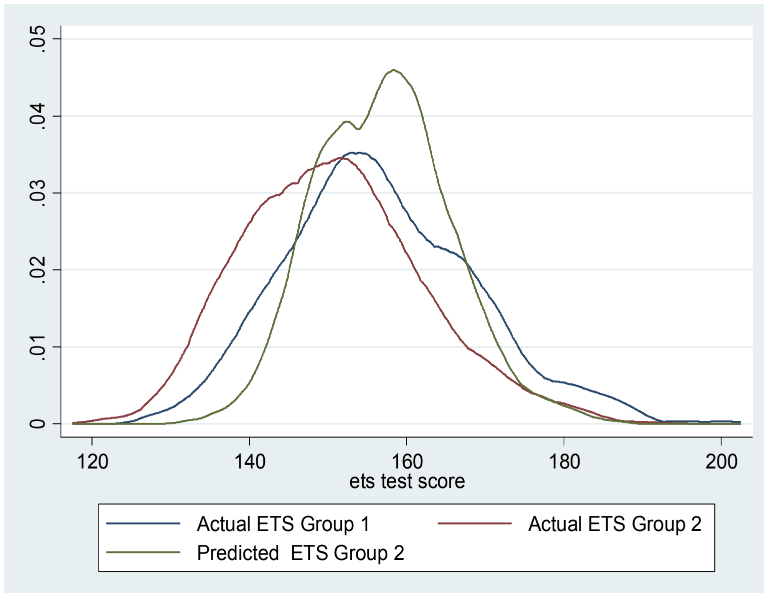
	MFT	GPA	CORE
Group_1	150.83*** (0.43)	3.06*** (0.02)	2.99*** (0.02)
Group_2	155.89*** (0.68)	2.92*** (0.03)	2.79*** (0.04)
Difference	-5.06*** (0.81)	0.13*** (0.03)	0.19*** (0.04)
Endowments	1.12 (0.73)	0.09*** (0.03)	0.12*** (0.04)
Coefficients	-5.76*** (0.72)	0.16*** (0.02)	0.14*** (0.03)
Interaction	-0.41 (0.60)	-0.12*** (0.02)	-0.06*** (0.03)

Column 1 of Table 8 lists the decomposition of MFT scores. Changes in student characteristics would have led to a 1.1-point increase in scores even though actual scores fell by 5 points. Changes in student characteristics lower MFT scores by 5.76 points which completely offsets the actual reduction in test scores, and 62% of the gain from the entitlement effect. The simultaneous changes in characteristics and coefficients is small and imprecisely estimated in the MFT decomposition as it represents what remains after accounting for both the entitlement and coefficient effects. The decomposition of GPA and CORE in columns 2 and 3 respectively reports an increase in grades, which is larger for the common CORE. Both endowment and coefficient effects boost grades, and the coefficient effect is 78% and 16% larger than the endowment effect for GPA and Core respectively. The combined increase from both sources results in a larger increase in grades than is actually reported during the period, which implies a negative interaction effect for the grade outcomes.

For both test scores and grades, the endowment effect would have led to increases in each, which is relatively large and significant for each grade and of modest size and just below the conventional significant levels for test scores. The divergence, hence, comes for the sharply different impact the characteristics have on the internal and external assessment measures as the coefficient estimates fall for the external test scores yet increase for both the GPA and CORE.

FIGURE 1

PREDICTED VERSUS ACTUAL ETS SCORES



The Blinder- Oaxaca decomposition accounts for changes at the mean but it is silent about other points of the distribution. Kernel densities provide clear visual evidence of the changes in student performance over the 20-year period. Figure 1 charts the MFT scores of each group along with a counterfactual distribution which is an estimated grade distribution based on the characteristics of Group 2 and the coefficients of Group 1. Test scores were higher in the early period at all points on the distribution. The difference in test scores is larger at the lower end of the distribution as the proportion of low MFT score grew over time. Modal differences between the groups are small compared to the fraction beyond one standard deviation from the mean. Predictions reduce the variances in the density as unobservable characteristics play a larger role in determining scores at the lower end of the distribution.

TABLE 9
THE DISTRIBUTION OF TEST SCORES AND GRADES BY COHORT: ACTUAL AND PREDICTED

Percentile

	10th	25th	50th	75th	90th
Panel A- ETS					
Group 1	142	149	155	161	171
Group 2	137	143	150	158	166
Predicted - 2	146.5	150.6	157.2	162.6	168.1
Panel B - GPA					
Group 1	2.37	2.59	2.92	3.30	3.63
Group 2	2.46	2.71	3.04	3.44	3.71
Predicted - 2	2.52	2.67	2.84	3.01	
Panel C - CORE					
Group 1	2.17	2.33	2.75	3.30	3.75
Group 2	2.22	2.54	3.00	3.50	3.80
Predicted - 2	2.34	2.52	2.78	3.04	3.32

The predicted scores are concentrated over higher values than the densities of actual scores suggesting lower performance of Group 2 at all point of the distribution and not just at the mean noted in Table 9. Predicted scores also dominates the actual scores for Group 1 implying that based on observables, test scores should have improved. For a more concrete comparison, Panel A of Table 9 presents the actual distribution of scores for each group along with the predicted score for Group 2 at various points of the distribution. Scores have declined at all points of the distribution, although the gap is widest below the median. Based on the characteristics of Group 2 and the regression coefficients for Group 1, we would predict significantly higher MFT scores not only for Group 2, but also higher than the actual for Group 1. Clearly, changes in characteristics over time would have at least maintained scores.

FIGURE 2
PREDICTED VERSUS ACTUAL GPA

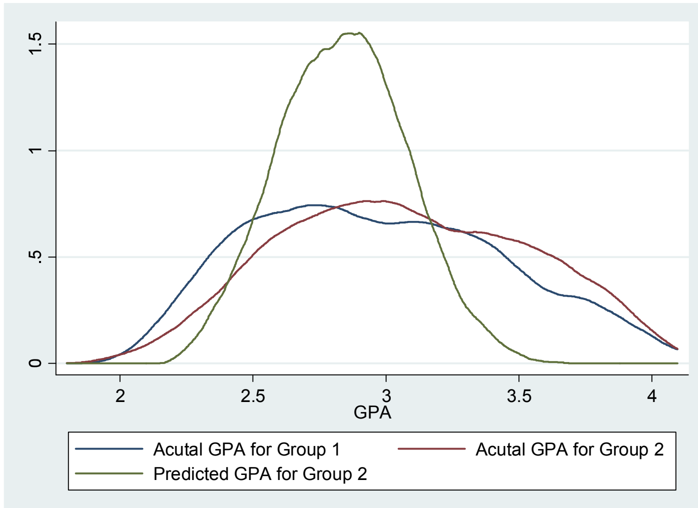
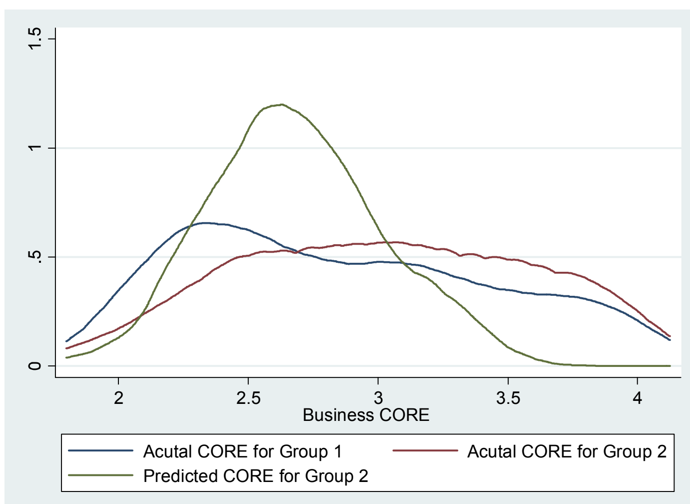


FIGURE 3
PREDICTED VERSUS ACTUAL CORE GPA



Figures 2 and 3 provide the same comparisons for GPA and Core grades that Figure 1 did for the MFT. Unlike the density of test scores, the density of both GPA and CORE has drifted higher. While grade inflation has occurred across many college campuses, the results here are within one major, Business, and the CORE shows changes within the same set of courses[7]. The kernel density of CORE grades has shifted further towards higher grades than has the density of overall GPA. The predicted GPA and CORE grades, based on Group 1 coefficients and Group 2 characteristics, are much more concentrated, but unlike for predicted test score, the largest reduction comes in the upper part of the density. Omitted characteristics are a larger part of the prediction error in the upper part of the distribution rather than in the lower part of the distribution as it was in test scores.

Panels B and C in Table 9 report the actual distribution and predicted distribution of both GPA and CORE, respectively, at the same points of the distribution as Panel A. Panel B shows that the median GPA grew by about .12 points and slightly larger increases are found at the 25th and 75th percentiles. The gap between the counterfactual predicted grade and the actual grade widens over the distribution from .04 grade points at the 25th percentile to .20 points at the median and reaching .43 grade points at the 75th percentile. Grades in the common CORE increased sharply between the 25th and 75th percentiles and were one-quarter of a letter grade higher at the median. The gap between the predicted and actual CORE grade is even wider than the gap between the predicted and actual GDP. The difference between the actual and predicted rises from one-quarter of a letter grade at the median to nearly one-half of a letter grade at the 75th percentile. Hence, within the same courses, it was substantially easier to get an A at the end of the study period.

DISCUSSION

This paper compares internal and external assessments in one program at one institution over a 20-year period in which grades have risen and test scores declined. Our results mirror this change as the reasonably close relationship between grades and test scores in the early part of the period has given way to a much weaker relationship later. Different regression models at the end of the period only explain between one-third and one-half of the variation in test scores as they did at the beginning of the period. Despite changes in student characteristics that should have increased test scores based on the regression estimates for test-takers between 1995 and 2001, sharply reduced coefficients in the later period dropped test scores by over 5 points. As both overall GPA and CORE grades have increased over time, changes in both student characteristics and changes in coefficients led to higher grades.

Higher grades are not surprising as grades have been increasing on college campuses for five decades. Sabot and Wasserman (1991) tie grade inflation to shift from low-grading majors in math and the sciences to higher grading majors in social sciences and the humanities. The grades reported here are within the business major although there are differences based on the concentrations within business. Furthermore, we find that grades have increased more with the same core courses than they have in the entire business field that leave little doubt about improving grades with not only major but in a given set of courses. The coincidentally declining test scores provide even stronger supports for grade inflation as it shows that improved grades are not due to better performance relative to the earlier period, unless the courses are testing different material than is on the standardized test. More detailed analysis of grade inflation is comprised by the lack of suitable data as all studies used only data from one institution owing to the private nature of college records. The results from all the studies ~~all~~ report grade inflation at the institution considered.

There is no lack of speculation about the reasons for higher grades although many can neither be verified nor refuted. Achen and Courant (2009) see grades as part of an intra-university economy that determine within university resource allocation. Professors have traditionally preferred small classes populated by talented and highly motivated students while administrators desire larger class sizes that they put under the heading of faculty productivity. Achen and Courant (2009) maintain that grades will be lower in required courses where the grade-elasticity of demand is low and higher in courses that use interpretative grading methods that give students more ability to contest low grades, which professors want to avoid. Grades are also higher in courses taught by non-tenure stream and untenured faculty as higher grades provide some benefit in the form of better student evaluations[8]. Grades also play a role in within university resource allocation and the role will be stronger when resources follow student enrollment. Departments compete with one another for students and may adjust grading standards if reduced enrollment threatened faculty lines and the opportunity to enjoy increased income from additional teaching. Achen and Courant (2009) argue that grades do not occur in a vacuum and that higher grades by one professor and in one department will provide incentive to others to follow. While grades are the purview of the professor, community standards require that one cannot have a criterion that it too much at variance with his or her colleagues. The five-decade long increase in grades has altered the standards and confronting the established expectations would lead to substantial resistance and lead to unpleasant workplace and perhaps threaten continued employment.

On the other hand, grade inflation reduces the information that is given to students, other institutions and perspective employers. Students select courses and major based on grades and grade inflation may lead students to courses for which they have no real comparative advantage. Higher grades also reduce the signal to post-college users of the information whether its employers or graduate admission officers. These users will quickly find other criteria to evaluate applications such as test scores or offer employment contracts that tie compensation to job performance.

Policy recommendations are not the concern of this investigation. We show a growing conflict between the internal and external assessment criteria at this institution. Outcomes in college and on tests represent an intermediate step for most students as they are more concerned about post-college employment. Future research could address how different assessment outcomes are related to post-college experience.

ENDNOTES

1. ETS reports that nationally, the scores fell by 4 points between 1995-2014
2. ETS reports, between September 2013 and May of 2015, 562 institutions gave the test to over 67,000 student
3. Prior to test administration, authors asked students about the degree to which they were motivated to take the MFAT-B and the extent to which the maintenance of AACSB accreditation was important to them
4. This difference is approximately equal to the adjustment that the College Board for students with scores between 400 and 450. See http://research.collegeboard.org/programs/sat/data/equivalence
5. The adjustment is the mean differences in the average score at this institution. Individual adjustments are higher in both math and verbal for students with scores in the lower end of the distribution
6. We utilize the Becker and Salemi (1977) general model of learning to estimate a specific functional learning function
7. Sabot and Wakeman-Lind (1991) find that changes in the distribution of students over majors have contributed to grade inflation
8. Rosovsky and Hartley (2002) show that the increase in grades corresponds to an increase use of non-tenure stream faculty and Johnson (2003) finds that student teaching evaluations improve with grades.

REFERENCES

Achen, Alexandra C and Paul Courant (2009), What Grades Are Made Of, *Journal of Economic Perspectives*, 23 (3), 77-97

Allen, J. S., & Bycio, P. (1997). An evaluation of the educational testing service major field achievement test in business. *Journal of Accounting Education*, *15*, 503–514.

Bagamery, B., Lasik, J., & Nixon, D. (2005). Determinants of success on the ETS Business Major Field Exam for students in an undergraduate multisite regional university business program. *Journal of Education for Business*, 81, 55-63.

Becker W. E., & Salemi, M. K. (1977). The learning and cost effectiveness of AVT supplemented instruction: Specification of learning models. *The Journal of Economic Education*, *8*, 77–92.

Betts, R. J., & Morell, D. (1999). The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects. *The Journal of Human Resources*, 34 (2), 268-293.

Bielinska-Kwapisz et. al (2012). Is Higher Better? Determinants and Comparisons of Performance on the Major Field Test in Business. *Journal of Education for Business,* 87, 159–169.

Blinder, A.S. (1973), Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources* 8, 436-455.

Bush, H. F., et al., (2008). Using The Major Field Test-Business As An Assessment Tool And Impetus For Program Improvement: Fifteen Years Of Experience At Virginia Military Institute. *Journal of College Teaching & Learning,* 5(2), 75-88.

Bycio, P., & Allen, J. S. (2007). Factors associated with performance on the educational testing service major field achievement test in business. *Journal of Education for Business*, *82*, 196–201.

Contreras, S., Badua, F., Chen, J.S., & Adrian, M. (2010). Documenting and Explaining Major Field Test Results Among Undergraduate Students. *The Journal of Education for Business*, 86(2), 64-70.

DiNardo, J., Fortin, M. N., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of wages, 1973-1992: A Semi parametric Approach. *Econometrica,* 64(5), 1001-1044.

Educational Testing Service. (2014). Major Field Tests: Business. Retrieved from https://www.ets.org/s/mft/pdf/mft_testdesc_business.pdf

Herring III, H.C. and Douglas, I.C. (1992). Outcomes assessment of accounting majors. *Issues in Accounting Education*. 1(7), 1-16.

Johnson, Valen E. (2003), Grade Inflation: A crisis in College Education, New York: Springer-Verlag

Mirchandani, D., Lynch, R., & Hamilton, D. (2001). Using the ETS Major Field Test in Business: Implications for assessment. *Journal of Education for Business*, 77, 51-56.

Oaxaca, Ronald L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 14 (3): 693–709.

Rosovsky, Henry and Matthew Hartley, (2002), Education and the Academy: Are We Doing the Right Thing? Cambridge MA, American Academy of Arts and Sciences

Sabot, Richard and J. Wakemann-Linn. (1991) "Grade Inflation and Course Choice", *Journal of Economic Perspectives*, (Winter 1991), 159-170.

Terry, N., Mills, L., & Sollosy, M. (2008). Student grade motivation as a determinant of performance on the business Major Field ETS exam. *Journal of College Teaching & Learning*, *5*(7), 27–32.

Wise, D. (1975). Academic Achievement and Job Performance. *American Economic Review*, 65, 350-366.