Journal of the Scholarship of Teaching and Learning, Vol. 12, No. 4, December 2012, pp. 43 – 57.

Measuring learning through cross sectional testing

Steve Lovett[1] and Jennie Johnson[2]

Abstract: The measurement of student learning is becoming increasingly important in U.S. higher education. One way to measure learning is through longitudinal testing, but this becomes especially difficult when applied to cumulative learning within programs in situations of low persistence. In particular, many Hispanic Serving Institutions (HSIs) find themselves in such situations. Cross sectional testing is a pragmatic alternative, so long as maturity and selection effects can be estimated. The purpose of this paper is to demonstrate the utility and mechanics of measuring learning through cross sectional testing.

Keywords: assessment, measurement, student learning, maturity, selection.

The measurement of student learning is fundamental to the scholarship of teaching and learning. Pace (2011) recently wrote:

> "...assessment is at the core of the entire SoTL enterprise. It is difficult to imagine a robust scholarship of teaching and learning unless our work is cumulative and built on previous research and unless there is a means to systematically evaluate the validity of claims being made about student learning" (p. 107).

Assessment, which begins with the measurement of learning, is becoming increasingly necessary for pragmatic reasons as well. Many parties, from prospective students to employers to governments, are increasingly demanding simple, quantitative performance measures from colleges and universities (Archibald & Feldman, 2008; Burke, 2002; Martell & Calderon, 2005), and accrediting bodies including the Association to Advance Collegiate Schools of Business (AACSB) and the European Quality Improvement System (EQUIS), are shifting their emphasis from input measures (number of faculty with terminal degrees, etc.) to outcome measures, or the extent to which students have achieved educational goals (Mundhenk, 2005; Rubin & Martell, 2009). The purpose of this paper is to demonstrate a simple, pragmatic procedure by which student learning may be measured through cross sectional testing. The focus is on cumulative learning within programs, not within individual courses. The sample with which we demonstrate the procedure comes from the business school of a Hispanic Serving Institution (HSI).

I. Measuring Learning.

The two most important student outcome measures are persistence – students should finish their programs – and learning – students should gain knowledge, skills or abilities while in their programs. The importance of these two outcomes has been recognized in recent academic research. In a literature search, Robbins, Lauver, Le, Davis, Langley, and Carlstrom (2004) found 408 studies reporting on at least one of these outcome measures between 1984 and 2003.

[1] School of Business, The University of Texas at Brownsville, 80 Fort Brown, EDBC. 1.434 Brownsville, TX, 78520, steve.lovett@utb.edu
[2] School of Business, The University of Texas at Brownsville, 80 Fort Brown, EDBC 1.436, Brownsville, TX,78520, jennie.johnson@utb.edu

Of the two, however, persistence is easier to measure. A student enrolled in one semester either enrolls the next semester or does not, and either graduates within some time frame or does not. Furthermore, the total reenrollment or graduation rate for a program is a valid measure of overall persistence within the program. The temptation for some institutions, therefore, may be to focus on the easily measured outcome – persistence – and to neglect the hard-to-measure outcome – learning. However such an unbalanced set of priorities is not to our advantage as a society. Archibald and Feldman (2008) wrote:

> "...not every strategy a university might design to increase its graduation rate would be a good educational decision. Universities could achieve a higher graduation rate by lowering curricular standards or by encouraging more grade inflation. And any institution could surely achieve higher graduation rates by restricting access to students who are sure bets to graduate. Raising graduation rates in these last two ways is not socially useful since it would weaken the country's commitment to high quality programs and broad based access" (p. 81).

What is needed to maintain a balance between the two outcome measures is a pragmatic means of measuring learning within programs. In this paper we will not offer the reader an instrument by which to measure learning. The content of such an instrument, while obviously a critical issue for any program, is outside the scope of this paper. Rather, this is a method paper – our purpose is to demonstrate a simple, pragmatic procedure by which learning may be measured through cross sectional testing.

This paper will provide the reader with an example of how to measure learning through cross sectional testing in a case-study format. We continue with seven more sections. The next section compares the indirect and direct methods of measuring learning. For example, one useful direct method is testing. The third section continues with a comparison of longitudinal and cross sectional testing. The fourth section explains why maturity and selection effects must be accounted for to measure net gains in learning. The fifth section describes the sample used for this demonstration – freshmen and seniors in a business program at a Hispanic Serving Institution (HSI) – and the sixth, the procedure used to estimated net learning gains. The final sections are a discussion with suggestions for improving and expanding the procedure, and a conclusion.

II. Indirect versus Direct Measurement.

A key step in managing any outcome is its measurement, and therefore all institutions have an interest in measuring student learning within their programs. Learning can be measured either directly or indirectly. The direct approach (used in this paper) is based on a *demonstration* of learning. The indirect approach is based on an *opinion* regarding learning, and this opinion may come from employers, alumni, teachers or even the students themselves (Martell & Calderon, 2005). However, the indirect approach has its shortcomings.

The simplest form of the indirect approach is self-assessment: ask the students how much they have learned. Perhaps because it is so simple, this method is commonly used, even in formal research. Sitzmann, Ely, Brown and Bauer (2010) found 166 recent studies in which self-assessment was used as a measure of learning. However, not all self-assessments are accurate. For example, Kruger and Dunning (1999) described a series of experiments that demonstrated that low-performers systematically overestimated their performances, and Clauss and Geedey (2010) found that students were reasonably good at self-assessing their knowledge or recall of facts, but much less able to self-assess their comprehension or their ability to apply knowledge. Sitzmann et al. (2010) performed a meta-analysis and found only weak correlations between self-assessments and learning. They concluded that "… self-assessed knowledge is generally more useful as an indicator of how learners feel about a course than as an indicator of how much they learned from it" (p. 180).

Another convenient measure of learning is the student's Grade Point Average (GPA). However, a course grade is ultimately an instructor's opinion, and different instructors often grade very differently, to the point that a student's grade in any given course falls far short of being a reliable demonstration of learning. Pace (2011) wrote "… procedures for determining grades are generally shrouded in mystery and rest upon processes that may be perfectly legitimate for classroom teaching but do not provide a firm foundation for a systematic exploration of teaching and learning" (p. 108).

Direct measures of student learning are to be preferred and accrediting bodies are beginning to emphasize these. For example, the Association to Advance Collegiate Schools of Business (AACSB) emphasizes direct measurement of learning in the new accreditation standards adopted in 2003, and this may be the most significant change from the prior standards (Thompson, 2004). AASCB acknowledges on their website that indirect measures may have some value, but states clearly that "Such indirect measures, however, cannot replace direct assessment of student performance" (AACSB, 2012). Martell and Calderon (2005) were even more blunt in their rejection of indirect methods of assessment, writing: "… we advise deans to forget about surveys and other indirect measures when thinking about assessment [because they have] very little evidentiary value for assessment of student learning" (p. 223). Therefore, this paper focuses on the direct approach, based on a *demonstration* of learning.

III. Longitudinal versus Cross Sectional Testing.

While students may demonstrate learning is a variety of ways, one method that gives the researcher significant control is through testing. If the students graduating from one program score higher on a test than those from another on the same test, it is reasonable to claim that the overall performance of the first program is better. The problem, however, is that the students in the first program might have been better prepared initially than those in the second, and a solution may be "before and after" or "value-added" testing – if students could be tested upon entering their programs and then again upon completion, the "gain" or the average difference in scores could be taken as a measure of learning. Of course, the students could also be tested at the beginning and end of a semester in order to evaluate learning within a course, or even several times during a semester (see, for example, Dellwo, 2010), but the emphasis in this paper is on the longer time period and on measuring learning within programs.

The preceding paragraph describes longitudinal testing – the same students are tested at the beginning and end of their programs. Adams and Schvaneveldt (1985) wrote "Advantages of this approach are that one observes or studies the same issue and the same people or events over a long period of time … [so] one can have greater control and ultimately more precise measurement with a longitudinal design" (pp. 115-116). However, they later added "A problem associated with the longitudinal design is the expense of following a population over time" (p. 116).

Rather than longitudinal testing, it would be simpler and less expensive to compare cohorts. For example, the entering class in a given year could be compared to the graduating class in the same year. This describes cross sectional testing. Adams and Schvaneveldt (1985) wrote “The cross sectional approach is obviously not as useful as the longitudinal design to assess change or development, but a number of inferences about change can be properly assessed within the constraints of this approach ... [and] ... for many research problems, the advantages of less time, fewer resources, larger samples, a large array of variables, and the versatility of the cross sectional approach indicate its central utility in social science research” (p. 115).

Cross sectional research is not without its critics. Seifert, Pascarella, Erkel and Goodman (2010) analyzed a large data set from nineteen U.S. institutions, first using a cross sectional method and then using a longitudinal method, and found significant differences between the results of the two methods. They concluded that that “... longitudinal pretest-posttest designs are the best way to estimate what students are learning … ” (p. 13), and they referred to longitudinal testing as the “gold standard” (p. 14) in assessment of student learning. However, they also stated that “We clearly recognize that longitudinal, pretest-posttest designs put a greater burden on institutional researchers in terms of time, effort and resources than do cross sectional studies” (p. 14), and this last point is especially relevant to institutions with low persistence.

When persistence is low only a minority of the beginning students may actually graduate from a program, and many of those graduating may be transfer students who began in another program, making the problem of “tracking” students from entrance to graduation extremely difficult. Unfortunately, the low persistence rate scenario appears to be the more common within U.S. higher education. ACT (2010) reported that first to second year retention rates in four-year public institutions recently averaged only about 68%, with only about 43% of students completing their degrees within six years. Furthermore, it is the institutions with fewer resources that suffer the most from low persistence. Gansemer-Topf and Schuh (2006), in a study of 466 American institutions, found that wealthier and more selective institutions had far better retention than poorer or less selective institutions, and in fact reported that “Institutional expenditures and institutional selectivity explained over 60% of the variance in retention and graduation rates” (p. 622). The authors of this paper are particularly aware of the difficulties of measuring learning in situations of low persistence. We work in the business school of a Hispanic Serving Institution (HSI), and we have been frustrated in the past by the problems involved in tracking large numbers of students from entry to graduation. According to the U.S. Census Bureau (2010), U.S. Department of Education (2011), and U.S. Department of Labor (2011), persistence is a particular problem in HSIs.

Given low retention rates, tracking individual students through a program for longitudinal testing becomes prohibitively expensive for many poorer or less selective institutions. Therefore, while longitudinal testing may be preferred, as a pragmatic matter cross sectional testing is less burdensome because it eliminates the need to track individual students, and is therefore more likely to be realized at most institutions. Furthermore, cross sectional testing produces results quickly, while a researcher using longitudinal testing must wait for students to complete their programs before getting results.

IV. Maturity and Selection Effects.

But at least two difficulties must be considered in cross sectional testing. The first is also a consideration in longitudinal testing; there may be a “maturity” effect. Graduating students are older than incoming students and so may have gained knowledge or skills independently of their university experience. For example, students gain vocabulary by watching television, and gain math skills by balancing their checkbooks. Pascarella and Terenzini (2005) recognized this difficulty, writing “It is one thing to conclude that increases in subject matter knowledge and academic skills occur *during college*. It is quite another to conclude that these increases occur *because of college*” (p. 70). We must therefore be aware that knowledge and skills can be gained by students outside of the program whose effectiveness we are trying to measure.

The second difficulty is unique to cross sectional testing. There may be a “selection” effect – poorly performing students tend to drop out of the program, while high performing students persist. Tinto’s (1975, 1993) model of student departure begins with pre-entry attributes such as skills, abilities and prior schooling – some students are more ready for college than others, and those who are less ready are less likely to perform well and therefore more likely to drop out. This relationship between readiness, performance and retention has documented by researchers such as Allen, Robbins, Casillas and Oh (2008), and the resulting selection effect will tend to raise the average performance of those remaining in the program, creating a false appearance of learning with the program.

Finally, note that maturity and selection are separate effects – that students learn outside of college is one thing, and that less prepared students are more likely to drop out is another. Therefore, both must be accounted for when attempting to measure learning within a program.

V. Sample.

The subjects of this study were two groups of students attending the business school of an open enrollment, Hispanic Servicing Institution (HSI). The U.S. Department of Education (2011) defines an HSI as a public or private non-for-profit college or university that has a student population of at least 25% Hispanic and serves a higher portion of low to middle income students than their peer universities. Research shows that Hispanics have fallen behind in educational attainment/program completion compared to white student groups as well as other cultural/ethnic groups (Alon, Domina & Tienda, 2010).

One group was comprised of freshmen in the introductory business course, and the other of graduating seniors in the program’s capstone course. Both groups coincidentally numbered sixty-three students. Near the end of the courses, both groups were given identical tests of sixty multiple choice questions which accounted for ten percent of the final course grade. The testing was therefore “course-embedded,” or part of a course, rather than an entrance or graduation requirement. However, note that the basic techniques demonstrated in this paper would apply equally well to entrance/graduation testing.

There were four categories of questions, one for each of the three subject matter areas emphasized in the program – management, marketing and accounting/finance – and quantitative questions. Three demographic questions were also asked: the student’s age (coded as a continuous variable), whether English was the student’s first language (dichotomous variable), and whether the student was male or female (dichotomous variable).

Information as to the students’ readiness for college was obtained from the school’s admissions center. All students, whether first time freshmen or transfers, are required to take reading, writing and mathematics tests before enrolling in classes. The scores from these three tests were standardized and the average of the standardized scores will be referred to below as a “readiness composite” or simply “readiness.” Similar data is available at most institutions.

Caison (2007) noted that “...institutions do routinely collect a broad array of information on their students’ backgrounds, socioeconomic status, academic progress, and, in many cases, their academic goals and social involvement” (p. 436), and found that this archival information was more useful in predicting retention than the information from a survey instrument.

Table 1 below shows descriptive statistics from the sample. The bottom of the table shows that the seniors did outperform the freshmen on all four outcome scores. The gains in management, marketing and accounting/finance were 0.64, 0.43, and 0.79 standard deviations. Pascarella and Terenzini’s (2005) “best estimate” (p. 66) of typical freshmen to senior gains for subject matter knowledge, taken from a synthesis of 17 previously published studies, is 0.87 standard deviations. These results show gains of somewhat less than that. The gains in the quantitative section were 0.90 standard deviations, greater than Pascarella and Terenzini’s (2005) estimate of 0.24 standard deviations.

Table 1. Descriptive statistics.

	Freshman (n=63)	Seniors (n=63)	Overall stddev	Gain in stddev
1. Average age	19.6	27.7	--	--
2. Readiness composite	-.357	.353	--	--
3. % with English as a first language	47.5	43.5	--	--
4. % female	44.4	49.2	--	--
% correct				
5. Management	69.0	78.6	15.0	0.64
6. Marketing	69.1	76.0	16.2	0.43
7. Acct/finance	67.7	81.5	17.4	0.79
8. Quantitative	57.7	78.4	22.9	0.90

The demographics shown at the top of Table 1 indicate a possible maturity effect because the seniors were older than the freshmen. Also, the seniors’ readiness composite was higher than the freshmen’s, indicating a possible selection effect. Both groups were relatively similar in the percent of students with English as their first language and percent women.

Table 2 below displays correlations between the variables. Age correlated significantly with three of the four outcome scores, indicating a likely maturity effect. The readiness composite correlated even more highly with all four outcome scores, indicating a likely selection effect. Also, note that age was statistically significantly correlated with the readiness composite. Because the analysis below is based on regression with these two as independent variables, this indicates potential problems with multi-colinearity. None-the-less, because the two represent separate effects (see previous discussion), both will be used below. English as a first language and percent female did not correlate significantly with any of the outcome scores, and as Table 1 shows there was little difference between the freshmen and seniors in these variables. Therefore, these last two are not used below.

Table 2. Correlations.

	1.	2.	3.	4.	5.	6.	7.
1. Age							
2. Readiness composite	.27*						
3. English first language	.09	.10					
4. Female	.03	.09	.10				
5. Management score	.20*	.39*	.00	-.00			
6. Marketing score	-.02	.30*	-.04	-.09	.47*		
7. Acct/finance score	.24*	.39*	.04	.11	.50*	.58*	
8. Quantitative score	.21*	.43*	.04	-.17	.51*	.36*	.38*

* correlation is statistically significant at the 0.05 level (2-tailed)

VI. Procedure.

Equation 1 will serve as a starting point for cross sectional testing.

(1) Gain = Average Senior Score – Average Freshman Score

Program gains independent of maturity and selection effects will be referred to as the "net gain" of the program, as summarized in equation 2.

(2) Net Gain = Gain – Maturity Effect – Selection Effect

In this study two evaluation items were sought. The first was an assurance that some positive net gains were being realized in each subject matter area. The second was a point estimate of the net gains in each subject matter area. That positive gains were being realized was verified by using a series of four regressions, each with one of the outcome scores as the dependent variable, and with age, readiness, and a dichotomous variable indicating senior status as independent variables.

(3) $\text{score} = b_0 + b_1(\text{age}) + b_2(\text{readiness}) + b_3(\text{senior})$

The results are shown in Table 3. In the cases of marketing, accounting/finance and quantitative skills, the results were encouraging – the beta value of senior status was positive and statistically significant at $p<0.05$. Therefore, even after age and readiness were considered, the seniors still scored significantly higher than the freshmen. In the case of management, however, the beta value of senior status was positive but not statistically significant.

However, these regressions simply show whether learning was statistically significantly different from zero, and a more appropriate goal is to maximize learning. We must therefore calculate a point estimate of learning or gains for each area. Based on the data from Table 1, we might naively calculate the gains in management, for example, as the simple difference between seniors and freshmen in management: 78.6 – 69.0 = 9.6 percentage points. However, this does not take into account the fact that the seniors were on average older (maturity effect) and more ready (selection effect) than the freshmen.

Table 3. Regressions.

	Dependent Variable							
	Management		Marketing		Acct/finance		Quantitative	
Variable	beta	sig.	beta	sig.	beta	sig.	beta	sig.
1. Constant	.706	.000	.821	.000	.694	.000	.672	.000
2. Age	.000	.916	-.006	.030	.000	.957	-.003	.325
3. Readiness	.058	.001	.051	.010	.056	.005	.083	.001
4. Senior	.052	.117	.078	.038	.096	.012	.174	.000
model r-square		.179		.127		.209		.274

We therefore need estimates of the maturity and selection effects. These were obtained from a series of four regressions, each with one of the outcome scores as the dependent variable, and with age and readiness as independent variables, but without senior status.

(4) $\text{score} = b_0 + b_1(\text{age}) + b_2(\text{readiness})$

The resulting b_1 and b_2 coefficients are unbiased estimators of the effects of age and readiness on students' scores without regard to whether the student was a senior or a freshman. The coefficients for the regression using the management score as the dependent variable, for example, are shown in Step A of Appendix 1.

Recall from Table 1 that the freshmen had an average age of 19.6 years and an average readiness composite of -.357. What would be the predicted score of such a group, without regard to senior status? Step B of Appendix 1 shows that we would predict a score of 69.8%, and that the equivalent predicted score of the seniors, with an average age of 27.7 and a readiness composite of .353, would be 76.2%.

The difference between these two figures – 6.4% – is the gain that can be explained through the maturity and selection effects, and this must be subtracted from the total gain to obtain the net gains of the program. Step C of Appendix 1 shows this calculation. The net gain in management was found to be only 3.2%, or 0.21 standard deviations, not the 9.6% calculated above. Repeating this methodology, Appendices 2 through 4 show that the net gain in marketing was 4.7% or 0.29 standard deviations, in accounting/finance it was 5.3% or 0.30 standard deviations, and in quantitative skills it was 9.3% or 0.41 standard deviations. The figures above are similar to Pascarella and Terenzini's (2005) estimate of 0.26 to 0.32 standard deviations for the net effect of attending college.

VII. Discussion.

First, note that the above analysis raises some concerns as to the sufficiency of the independent variables used. In particular, readiness (selection) appears to have had more powerful effects than age (maturity). Age was statistically significantly correlated with three of the four outcome variables, but readiness correlated even more strongly with all four outcome variables (see Table 2). Furthermore, when both age and readiness were included as independent variables in a regression equation, readiness was statistically significant while age was not. Now, this may be because selection is in fact the more powerful effect, however it also may be because selection was well measured by the "readiness composite" (calculated through admissions center data), while maturity was poorly measured by age.

In general, readiness and therefore selection is likely to be relatively easy to measure. For example, ability measures such as SAT or ACT scores, and early academic performance measures such as high school graduation rank or GPA have been shown to be reliable predictors of success in college (Harackiewicz, Barron, Tauer, & Elliot, 2002; Pascarella & Terenzini, 2005; Willingham, Lewis, Morgan, & Ramist, 1990), and most institutions have these or similar measures available for both freshmen and seniors. However, measuring maturity means estimating the value or applicability of life experiences to a subject matter, and this is likely to be quite difficult. The authors, for example, have witnessed some very thought provoking classroom debates between students with different backgrounds as to whose experiences made them better able to correctly understand or interpret a particular business case or dilemma. Age as an estimate of maturity is certainly easy to obtain – most students are quite willing to tell a researcher how old they are – but the effect of a year's passing on one student may be very different from its effect on another. Still, there is no doubt that some measure of maturity is needed because students learn outside of the classroom, so a better measurement of maturity is a useful topic for future research.

In any case, the results presented here would naturally direct a faculty's attention toward examining the management area more closely. This was the area in which the least gains were found, and therefore it might present the greatest opportunity for improvement. But again, note that there are two possible reasons why less gains might be found in management. First, it could be that students are in fact learning less in this area. However, it could also be that the test did not accurately measure what they did learn. The issue here is the content of the measurement instrument, or, in other words, the content validity of the dependent variable, and as noted in the introduction, this was outside the scope of this paper. A good source for guidance in this area, however, is Martell and Calderon (2005), who describe the creation of such an instrument as an on-going five-step process in which an entire faculty reflect on what they want their students to learn, define learning goals, measure student performance relative to these goals, report and discuss the results, and then make appropriate changes in teaching or curriculum.

Ideally, the results above might provide a stimulus for the management faculty to begin this process. If so, an initial benefit might be a clearer consensus on learning goals among the management faculty, which in turn might lead to a better measurement instrument. Of course, this has value in itself. Still, if the faculty do no more than this they will have limited themselves to what Vockell and Asher (1995) describe as level 1 research, or data collection. Level 2 research, which depends on level 1 research, means establishing cause-and-effect relationships. In this case the management faculty might experiment with new teaching techniques or curriculum in an effort to improve learning, which is the fifth step in the Martell and Calderon (2005) process mentioned above.

Another interesting topic is the application of the techniques described in this paper to other learning outcomes and other assessment methods. Kraiger, Ford and Salas (1993) and Kraiger (2002) describe three kinds of learning outcomes – cognitive, skill-based and affective – each of which can be assessed through a variety of methods. The dependent variable in this paper is a score on a multiple-choice test, so a cognitive outcome was measured through recognition testing. However, many institutions also seek to teach skills, such as critical thinking or writing, which may be assessed through observation or by reviewing work samples. If, for example, student writing samples could be reliably scored, these scores could take the place of the dependent variables used in this paper and net gains in writing skills could be estimated using the same method. Finally, institutions may seek affective outcomes such as positive attitudes or motivation, and, as Sitzmann et al. (2010) concluded, these may be accurately measured through self-reports. The techniques described in this paper, in which maturity and selection effects are estimated and subtracted from gains to find net gains, are applicable to any of these *so long as* reliable numeric scores can be assigned to student performances.

VIII. Conclusion.

Just as excellence in teaching and learning requires openness to different modes of delivery and instructional approaches, practitioners and researchers need to be open to various approaches to the measurement of learning. This paper previously discussed the pros and cons of longitudinal and cross sectional testing to measure learning. Longitudinal testing is useful, but less practical, especially in institutions with low persistence. Such institutions may benefit more from cross sectional testing for measuring learning. According to the U.S. Census Bureau (2010), U.S. Department of Education (2011), and U.S. Department of Labor (2011), persistence is a particular problem in Hispanic Serving Institutions (HSIs), which have lower graduation rates than non-HSIs. Of course, this will have an important impact on tomorrow's workforce due to the continued growth of the Hispanic population in the United States. Therefore, the cross sectional method of measuring learning demonstrated in this paper via case study has particular implications for programs in HSIs and other higher education institutions with similar challenges**.** Cross sectional testing can be used to estimate students' net gains or learning within a program, and the procedure demonstrated here is simple and well within the means of most institutions.

Appendices

Appendix 1. Net Gain in Management.

Step A - regression on management score.

Variable	beta	std.err	sig.
Constant	.683	.047	.000
Age	.002	.002	.230
Readiness	.068	.016	.000

model r-square .162

Step B - calculate predicted values based on age and readiness.

Freshmen	Seniors
.683	.683
+ (.002* 19.6)	+ (.002 * 27.7)
+ (.068 * -.357)	+ (.068 * .353)
-----------------	-----------------
.698	.762

Step C - calculate net gain.

(.786 - .690) - (.762 - .698) = .096-.064 = .032

= 0.21 std. dev.

Appendix 2. Net Gain in Marketing.

Step A - regression on marketing score.

Variable	beta	std.err.	sig.
Constant	.787	.053	.000
Age	-.003	.002	.228
Readiness	.065	.018	.000

model r-square .095

Step B - calculate predicted values based on age and readiness.

Freshmen	Seniors
.787	.787
+ (-.003 * 19.6)	+ (-.003 * 27.7)
+ (.065 * -.357)	+ (.065 * .353)
-----------------	-----------------
.705	.727

Step C - calculate net gain.

(.760- .691) - (.727 - .705) = .069 - .022 = .047

= 0.29 std. dev.

Appendix 3. Net Gain in Accounting/Finance.

Step A - regression on accounting/finance score.

Variable	beta	std.err.	sig.
Constant	.652	.055	.000
Age	.004	.002	.079
Readiness	.074	.019	.000

model r-square .166

Step B - calculate predicted values based on age and readiness.

Freshmen	Seniors
.652	.652
+ (.004* 19.6)	+ (.004 * 27.7)
+ (.074 * -.357)	+ (.074 * .353)
-----------------	-----------------
.704	.789

Step C - calculate net gain.

(.815 - .677) - (.789 - .704) = .138 - .085 = .053

= 0.30 std. dev.

Appendix 4. Net gain in quantitative score.

Step A - regression on quantitative score.

Variable	beta	std.err.	sig.
Constant	.596	.071	.000
Age	.004	.003	.219
Readiness	.115	.024	.000

model r-square .195

Step B - calculate predicted values based on age and readiness.

Freshmen	Seniors
.596	.596
+ (.004 * 19.6)	+ (.004 * 27.7)
+ (.115 * -.357)	+ (.115 * .353)
-----------------	-----------------
.633	.747

Step C - calculate net gain.

(.784- .577) - (.747 - .633) = .207 - .114 = .093

= 0.41 std. dev.

References

AACSB. (2012). www.aacsb.edu/accreditation/business/standards/aol/learning_goals. Retrieved June 4, 2012.

ACT, Inc. (2010). *National collegiate retention and persistence to degree rates.* Iowa City, IA.

Adams, G.R., & Schvaneveldt, J.D. (1985). *Understanding Research Methods.* White Plains, NY: Longman.

Allen, J., Robbins, S., Casillas, A., & Oh, I. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education, 49*, 647-664.

Alon, S., Domina, T., & Tienda, M. (2010). Stymied mobility or temporal lull? The puzzle of lagging Hispanics college degree attainment. *Social Forces, 88*(4), 1807-1832.

Archibald, R.B., & Feldman, D.H. (2008). Graduation rates and accountability: Regressions versus production frontiers. *Research in Higher Education, 49,* 80-100.

Burke, J.C. (2002). *Funding Public Colleges for Performance.* New York: Rockefeller Institute Press.

Caison, A.L. (2007). Analysis of institutionally specific retention research: A comparison between survey and institutional database methods. *Research in Higher Education, 48,* 435-451.

Clauss, J., & Geedey, K. (2010). Knowledge surveys: Students ability to self-assess. *Journal of the Scholarship of Teaching and Learning, 10*(2), 14-24.

Dellwo, D.R. (2010). Course assessment using multi-stage pre/post testing and the components of normalized change. *Journal of the Scholarship of Teaching and Learning, 10*(1), 55-67.

Gansemer-Topf, A., & Schuh, J. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education, 47*, 613-642.

Harackiewicz, J.M., Barron, K.E., Tauer, J.A., & Elliot, A.J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*, 562-575.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.) *Creating, Implementing, and Managing Effective Training and Development: State-of-the-Art Lessons for Practice* (pp. 331-376). San Francisco: Jossey-Bass.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311-328.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Martell, K., & Calderon, T. (2005). Assessment in business schools: What it is, where we are, and where we need to go now. In K. Martell and T. Calderon (Eds.) *Assessment of Student Learning in Business Schools: Best Practices Each Step of the* Way (pp. 1-26). Tallahassee: Association for Institutional Research.

Mundhenk, R. (2005). Assessment in the context of accreditation. In K. Martell and T. Calderon (Eds.) *Assessment of Student Learning in Business Schools: Best Practices Each Step of the Way* (pp. 27-42). Tallahassee: Association for Institutional Research.

Pace, D. (2011). Assessment in history: The case for "decoding" the discipline. *Journal of the Scholarship of Teaching and Learning, 11*(3), 107-119.

Pascarella, E.T., & Terenzini, P.T. (2005). *How College Affects Students: A Third Decade of Research.* San Francisco: Jossey-Bass.

Robbins, S., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130,* 261-288.

Rubin, R., & Martell, K. (2009). Assessment and accreditation in business schools. In S. Armstrong and C. Fukami (Eds.) *The SAGE Handbook of Management Learning, Education and* Development (pp. 364-384). Los Angeles: Sage.

Seifert, T.A., Pascarella, E.T., Erkel, S.I., & Goodman, K.M. (2010). The importance of longitudinal pretest-posttest designs in estimating college impact. *New Directions for Institutional Research,* Assessment Supplement, Winter, 5-16.

Sitzmann, T., Ely, K., Brown, K., & Bauer, K. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning and Education, 9*(2), 169-191.

Thompson, K. (2004). A conversation with Milton Blood: The new AACSB standards. *Academy of Management Learning and Education, 3*, 429-439.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45,* 89-125.

Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition* (2nd ed.). Chicago: University of Chicago Press.

United States Census Bureau. (2010). *Educational attainment in the United States: 2010 Detailed Tables.* Retrieved April 29, 2012 from https://www.census.gov/hhes/socdemo/education/data/cps/2010/tables.html

United States Department of Education. (2011). *Moving America forward: President Obama's agenda for the Latino community*, pp. 1-5. Retrieved April 29, 2012 from http://www.2.ed.gov/about/inits/list/hispanic-initiative/obama-agenda.pdf.

United States Department of Labor. (2011). *The Hispanic labor force*. Retrieved April 29, 2012 from http://www.dol.gov/_sec/media/reports/hispanicelaborforce.pdf.

Vockell, E. L., & Asher, J. W. (1995). *Educational Research* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Willingham, W.W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting College Grades: An Analysis of Institutional Trends over Two Decades.* Princeton, NJ: Educational Testing Service.