WHITE PAPER

Student Learning Assessment in Higher Education

Pedro Reyes, Associate Vice Chancellor for
Academic Planning and Assessment
The University of Texas System

TABLE OF CONTENTS

WHITE PAPER

Student Learning Assessment in Higher Education

February 2006

I. Policy Context

For many years, higher education leaders have avoided defining student learning as the key indicator to determine institutional excellence. Instead, they have, along with national publications, used a series of actuarial data as the indicator of success. The *U.S. News and World Report* annual issue on college and university quality practically has determined how institutions of higher education should be assessed. In doing so, the publication raters have used input measures such as entering class SAT scores, student–faculty ratios, reputational survey scores, and amount of dollars spent to support students (among others) to assess the quality of higher education institutions. Yet, scholars recently have questioned these approaches to assess institutional quality (Callan & Finney, 2000; Ewell, 2002), indicating that such inputs are only part of how quality should be defined. They advocate assessing student learning as the fundamental purpose of higher education.

Why is it Important to Redefine Quality Education in Higher Education and Use Student Learning as the Measure for Success?

Student learning is important because the public (taxpayers, legislators, governors, and parents, among others) wants to be assured that college students are receiving a quality education. The public wants to know if the investment in higher education is yielding outcomes consistent with expectations. That is, stakeholders want to know if students have acquired the skills, knowledge, and abilities they need when they graduate. Thus, the old way to assess quality (through accreditation, seat time, course grades, and graduation rates) is a thing of the past. New student learning assessment systems must be developed and implemented to satisfy the demand for accountability.

Many scholars agree that one of the most critical indicators of an institution's performance is its ability to improve the quality of student learning (Astin, 1991; Banta, 2002; Ewell, 2002; Pascarella, 2001). To accomplish this goal, institutional measures need to be established and linked to provide evidence of a positive impact on students (Ewell, 2002). Thus, investigating how out-of-class influences and in-class courses lead to better student learning outcomes is important. Equally important are measures of student level of proficiency in specific, core academic areas such as reading.

What Do We Know About Defining Quality and its Relationship to Student Learning Assessment?

Today's students desire a competitive level of knowledge and skills to do well in an ever-changing global market (Callan & Finney, 2002). Without student assessment, we cannot

say with any direct evidence if graduates are receiving the best education possible, given the costs and outcomes associated with earning a degree (Brunner, 1997). Quality education is hard to define and measure but is important, particularly to the public footing the bills (Klein, Kuh, Chun, Hamilton, & Shavelson, in press; Pascarella, 2001).

Most scholars believe past performance determines future outcomes (Klein et al., in press). However, there is debate over how much of a student's learning is based on classroom instruction. An additional area of interest is assessing what outside interactions and experiences impact learning. Shavelson and Huang (2003) claimed that gaining an understanding of how core abilities affect how knowledge is produced is consequently valuable and revealing in its attempts to understand how one learns and applies knowledge. They recommended using multiple tests with different approaches to measuring student outcomes. Further, understanding the definitions of the most basic level of learning to the most complex provides researchers with tools for better defining and choosing assessments best suited for their needs (Shavelson & Huang).

Assessments measuring "within-college experiences" provide a more revealing picture of a student's experience by measuring academic rigor, educational experiences and outcomes, and connections to faculty and staff (Pascarella, 2001, p. 19). We know an institution's role in student success is a direct indicator of its effectiveness and quality. Improving academic skills among undergraduates should trump the incessant drive to recruit the most academically competitive students and well-respected faculty.

What Are the Issues Affecting Student Learning Assessment?

Creating a valid sample given the number and complexity of student-centered factors is no easy task. Recruiting a small student sample can be a barrier to producing accurate, robust results if a large recruitment effort is not undertaken to ensure a strong yield for participation (Klein et al., in press). One strategy to increase participation and motivation on the part of the test-takers is to provide reasonable financial incentives (Klein et al.). Beyond attaining a large and random sample, longitudinal studies have inherent issues due to troubling student attrition and student ability to participate in periodic assessments due to time constraints set by classes and employment. Another tactic growing in popularity is the creation of a "capstone course" where the grade equals the outcome (Palomba, 2002, p. 207). Palomba described this assessment as embedded in a course required for one's degree.

II. Examples of Efforts to Assess Student Learning

In the 1970s, Alverno College and a few other institutions made significant contributions toward the understanding of assessment and its potential to change and improve the quality of higher education. Alverno College identified eight "outcome taxonomies" to measure, including communication, problem solving, analysis, and involvement in the contemporary world (Astin, 1991, p. 43). Faculty recommended these areas as those they valued and wished to assess. One significant outcome from this process was the faculty's desire to improve "teaching-learning methods" based on the data acquired (Astin, 1991, p. 43).

Northeast Missouri State University (NMSU) at Kirksville embarked on an assessment study during the 1970s that examined student background, campus impact on the student's life, strength of degrees earned, and the competitiveness of student graduates in their chosen careers (Krueger & Heisserer, 1987). Krueger and Heisserer found that the variety of assessments and questionnaires implemented at NMSU, the use of a database with longitudinal data, and the appropriate use of multivariate analysis made this institution an example of good assessment and decision-making practices. Some researchers also are focusing specifically on outcomes of underachievers and middle-of-the-road students, because these students represent the largest population that could gain from performing above the predicted learning outcomes modeled in regression analysis (Lingenfelter, 2003).

With the creation of conferences and professional collaborations, opportunities emerged to publicize and share academic papers and programs on assessment (Ewell, 2002). The first formal group was the American Association for Higher Education (AAHE) Assessment Forum, which began meeting annually in 1987 and offered a sense of collegiality to the few researchers in the field (Ewell, 2002). The Business-Higher Education Forum (BHEF, 2004), a national organization comprised of business and higher education leaders, published a report and recommendation in "Public Accountability for Student Learning in Higher Education" to forward their insights on excellence in education. BHEF has a strong interest in maintaining educational excellence as a means to satisfying workforce expectations, particularly since the group is comprised of nationally known industry representatives.

The emergence of literature and journals in the 1980s and early 1990s was led by the *New Directions for Institutional Research* series (Ewell, 2002). In 1989, *Assessment Update* became the one of the first refereed journals specifically marketed to the "faculty-practitioner" wanting to share methodologies and interesting findings. (Ewell, p. 14). The already established *Change* magazine published short essays in areas of assessment by emerging scholars also served an important role. Finally, the AAHE resource guides are a compilation of expanded pieces presented at the AAHE Assessment Forum, where experts in the field gather and share the latest research in assessment (Ewell).

III. Challenges in Implementing Learning Assessment

Student assessment in higher education provides the context for an evolving and growing profession engaged in studies that measure the value of postsecondary education in important and distinct ways. Over the years, three key reasons for measuring educational outcomes slowly and steadily have emerged: (a) to improve the quality of student learning, (b) to advance academic expectations and standards, and (c) to satisfy the public's growing appetite for accountability of its tax dollars.

Modern issues in assessment include finding valid comparison groups; incorporating a prompt, thoughtful response to findings; and addressing the need for student compensation as a tool to improve the sample validity and motivation of test-takers. Advancing educational success involves creating the right climate and support system for

student assessment (Peterson & Vaughan, 2002). Identifying peer groups provides opportunities for comparisons among cooperative institutions with distinct similarities yet potentially different assessment outcomes (Pike, 2002). According to Pike, national norm groups represent the larger population's performance on a standardized test and are desired as comparison groups, given the potential for better understanding of an institution's performance. The ability to use impartial methods for assessment provides a reliable and more valuable option for analysis when employing group comparisons.

A growing number of testing companies are providing "nationally normed groups for standardized tests," which enable institutional researchers to make more powerful and insightful conclusions based on performance of one group against a larger group more representative of the population (Gray, 2002). Similarly, peer institutions for qualitative assessments also should be included as a method for comparing an institution's performance against similar institutions on criterion-referenced measures (Pike, 2002). Although difficult to obtain, some scholars also recognize using self-selecting groups or non-controlled groupings as an analytical tool to support findings and explaining dissimilarities in inter-institutional experiences and outcomes (Pike). Once the analysis is complete and widely distributed, it is imperative to provide a mechanism for legitimate and timely responses to findings so that the entire college community feels compelled to consider the results in their decision-making processes (Pike).

Identifying strategies to conduct valid studies with a random sample plays an important role in implementing a sound assessment plan. Providing incentives to compensate test-takers for their time is becoming more financially challenging due to the number of students working and expecting reasonable compensation for their lost wages (Klein et al., in press).

Another key concern is student motivation to do well on tests, particularly given an undergraduate's status as a volunteer (Klein et al., in press). At the heart of sound assessment is recruiting a sufficiently large random sample to make valid conclusions and inferences (Pascarella, 2001). Issues of cost and time necessary to develop an assessment program and an integrated database are sensitive areas for publicly funded institutions, given their shrinking state-provided appropriations for current operations (Astin, 1991).

Assessment can be the catalyst for curricular improvement, but it is important to recognize, above all else, that any assessment tool has shortcomings. Beyond the struggle of finding the ideal instrument, the most critical challenge is institutionalizing assessment and making it central to the institution's commitment to excellence (Maki, 2002). Equally important is establishing policy and practices influenced predominately by findings and not by money or politics (Astin, 1977). Finally, the ultimate challenge is to create an institutional plan that defines assessments to be used and that pairs them with the key goals of the college (Shavelson & Huang, 2003).

IV. Potential Solutions

This section summarizes several assessment tools marketed to institutions of higher education. Creating university-wide involvement and utilizing an array of assessment

tools increase the likelihood of a more diverse student group's participation in a specific assessment (Klein et al., in press; Lingenfelter, 2003). These techniques also improve the possibility for analysis across fields and departments.

1. An example of a test frequently used to measure "selected academic skills…in core general education classes" is the CAAP by ACT (2005, para. 3). This paper-and-pencil test, created in 1992, has the flexibility to test one area or as many as six areas, depending on the student skills an institution wishes to assess. Test areas include writing skills, essay writing, reading, mathematics, science, and critical thinking.

2. The Collegiate Learning Assessment (CLA) developed by the Council for Aid to Education is a newer tool developed to measure critical thinking and analytical skills (Klein et al., in press). The CLA uses various types of performance and analytic writing tasks, all of which require open-ended responses. The assessment has no multiple-choice questions. The test has two sections: the performance task and the analytic writing task.

Performance tasks require students to use an integrated set of critical thinking, analytic reasoning, problem-solving, and written communication skills to answer several open-ended questions about a hypothetical, but realistic, situation. Students are provided with a document library for each task, which includes a range of information sources such as letters, newspaper articles, and diagrams to use in preparing their answers. All of the CLA performance tasks require students to present their ideas clearly and to justify the basis for their points of view.

Analytic writing tasks require students to write answers to two types of essay prompts: "Make-an-Argument" and "Critique-an-Argument." A "Make-an-Argument" question asks students to support or reject a position on a particular issue. A "Critique-an-Argument" question asks students to evaluate the validity of an argument made by someone else. Both tasks measure a student's ability to articulate complex ideas, examine claims and evidence, support ideas with relevant reasons and examples, sustain a coherent discussion, and use standard written English.

3. Student surveys such as the National Survey of Student Engagement (NSSE) are also useful and widely used by 4-year universities to measure student perceptions and attitudes about their college experience.

4. Another example of assessment is the Academic Profile, developed by Educational Testing Service (ETS), which was renamed in 2006 as The Measure of Academic Proficiency and Progress (MAPP). This test is a measure of college-level reading, mathematics, writing, and critical thinking in the context of the humanities, social sciences, and natural sciences. The MAPP test is designed for colleges and universities to assess their general education outcomes, so they may improve the quality of instruction and learning. It focuses on the academic *skills* developed through general education courses (ETS.ORG).

5. Lastly, College Base, a criterion-referenced exam, was created by the University of Missouri–Columbia to address proficiencies similar to the Academic Profile by testing various levels of challenge in core academic areas (Astin, 1991).

V. The University of Texas System Learning Assessment Initiative

Introduction

In 2004-05, The University of Texas System contracted with the RAND Corporation's Council for Aid to Education to conduct a Collegiate Learning Assessment (CLA) of each academic university within the UT System. The purpose of the assessment is to understand how well students are doing with respect to certain learning outcomes compared with their predicted performance and the performance of students at other undergraduate institutions. The assessment focuses on critical thinking, reasoning, problem solving, and writing, not on specific course-related knowledge. Tests are administered to all or a sample of an institution's freshmen and seniors and results are compared against those obtained from other similar institutions. Nationwide, a total of 120 institutions participated in the 2004-05 assessment.

This study helps answer several important questions. First, how well do the learning outcomes of students enrolled in UT System institutions compare to students from other institutions? Second, do students at UT System institutions, relative to students from other institutions, perform above, at, or below 'expected' levels based on their entering admissions test scores? Third, have the institutions added value as indicated by seniors showing levels of learning higher than expected relative to that expected of freshmen?

The UT System began administration of this assessment in 2004-05 to establish a baseline from which future progress can be measured. Results from this preliminary phase of assessment show that for all campuses with sufficient numbers of assessment participants, overall performance was at and, in some cases, above expected levels

The starting point for this assessment is that conceptions of university quality should be influenced by improvements in student learning. Although educational quality is often based upon such indirect measures as the test scores of entering students, opinion polls of experts, or available financial resources, the Collegiate Learning Assessment (CLA) bases its assessment on students' demonstrated abilities.

The CLA data are compared with the student's starting point. This study uses entrance examination scores as a measure of a student's "starting point" in college. Even thought the entrance exam scores are not a sufficient measure of preparation, the scores can be taken as a proxy for how well prepared a student is for college study. This is an important starting point, because a school whose students have very high entry credentials is limited in the value it can add, because the students are already near the top of the measuring scale. In the UT System, both UT Austin and UT Dallas have freshman classes with high entry credentials. By contrast, a school whose students have low entry credentials can add a great deal of value. Even if such students later score at only at the national average, their college attendance will have added substantial value. In the UT System, UT San Antonio is an example of this pattern.

The "expected" scores are statistical projections based upon the score a student would be expected to earn given entrance exam scores. These expected data can then be compared with students' actual scores.

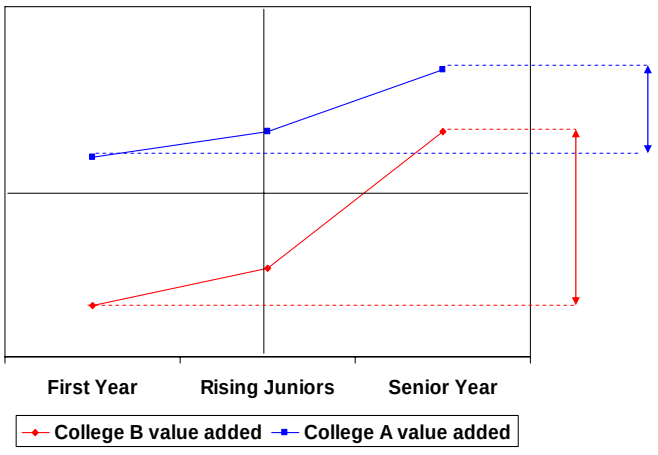
The CLA data for seniors are also compared with the CLA data for freshman students.

INSTITUTION – The primary unit of analysis is the institution. This means that the focus is on how the institution as a whole contributes to student development. Thus, we aggregate the information to understand better the institution's role in promoting learning.

VALUE ADDED – The CLA assessment focuses on the value added by colleges and universities. When institutional quality is based solely on students' scores on entrance examinations, there is no way to know what was learned after they matriculated; again, when student ability is only measured upon graduation, there is no way to determine the students' relative growth without knowing their starting point. It is only by comparing what students know when they start college with what they know when they finish that it is possible to assess the learning that actually occurred while in college.

COMPARISONS – This approach to assessment also allows for inter-institutional comparisons of overall value added. For example, the figure below shows how College A and College B added value to student learning both in terms of absolute scores and in terms of the difference in adding value given their respective student populations.

Figure I. Possible Comparisons of Institution's Value Added



What does the test measure?

The Collegiate Learning Assessment (CLA) uses various types of performance and analytical writing tasks, all of which require open-ended responses. These are no

multiple-choice questions. There are two sections to this test: 1) the performance task; and 2) the analytic writing task.

Performance tasks require students to use an integrated set of critical thinking, analytic reasoning, problem solving, and written communication skills to answer several open-ended questions about a hypothetical, but realistic, situation. Students are provided with a document library for each task, which includes a range of information sources such as letters, newspaper articles, and diagrams, to use in preparing their answers. All of the CLA performance tasks require students to present their ideas clearly, including justifying the basis for their points of view.

Analytic writing tasks require students to write answers to two types of essay prompts: "Make-an-Argument" and "Critique-an-Argument." A "Make-an-Argument" question asks students to support or reject a position on a particular issue. A "Critique-an-Argument" question asks students to evaluate the validity of an argument made by someone else. Both tasks measure a student's ability to articulate complex ideas, examine claims and evidence, support ideas with relevant reasons and examples, sustain a coherent discussion, and use standard written English.

How to interpret the scores

This study helps answer several important questions. First, how well do the learning outcomes of students enrolled in UT System institutions compare to students from other institutions? Second, do students at UT System institutions, relative to students from other institutions, perform above, at, or below 'expected' levels based on their entering admissions test scores? Third, have the institutions added value as indicated by seniors showing levels of learning higher than expected relative to that expected of freshmen?

To facilitate reporting results across institutions, the CLA scores were converted to the same scale (1 to 36) of measurement used to report ACT scores. The ACT scale has a mean of 20 and a standard deviation of 5. The CLA scale has the same properties. Roughly two-thirds of all students will score between 15 and 25. About one-sixth of all students will score below 15, and about one-sixth will score above 25. Caution should be used in interpreting relatively small differences (say, 20.5 versus 21) because those differences might not be statistically significant due to chance variation.

Key Findings

1. UT System freshmen perform on the CLA Tests as well as other institutions in the national sample.

Two tests are being reported: the performance task, which involves synthesizing and integrating materials to produce a document, and the analytic writing task. The analytic writing task requires students to write two essays, one that criticizes an argument and another one that makes an argument. The focus of the analytical writing test is on examinees' critical thinking and analytical writing skills, such as the ability to: 1) articulate complex ideas clearly and effectively; 2) examine claims and accompanying evidence; 3) support ideas with relevant reasons and examples; 4) sustain a well focused, coherent discussion; and 5) control the elements of Standard English.

For the nation as a whole, the expected score on the performance task scale, which is predicted from ACT scores, would be 22.5 for the Performance Task and 22.9 for the Analytic Writing task. The national expected scores are the same for each institution. The institutional expected scores vary because of differences in the ACT scores of students from each institution.

In summary, Table 1 shows the comparison of freshman CLA average scores for UT System institutions with the national study group sample. It shows that UT Dallas freshmen scored well above the average of the national sample in both the performance task and the analytic writing tests. This table also shows that UT Arlington, UT Permian Basin, UT San Antonio, UT El Paso and UT Pan American freshmen scored about the same as other freshmen in the national sample. UT Austin, UT Tyler, and UT Brownsville did not have enough freshmen in the sample for this analysis.

Table 1

University of Texas System
Freshman-Level CLA Scores by Institution

Institution	Measure	National Expected Score	Expected Institution Score	Actual Institution Score	National Comparison
Arlington	Performance Task	22.5	22.64	22.5	As expected
	Analytic Writing Task	22.9	22.81	22.0	As expected
Dallas	Performance Task	22.5	26.21	27.1	As expected
	Analytic Writing Task	22.9	26.29	26.7	As expected
El Paso	Performance Task	22.5	20.29	20.4	As expected
	Analytic Writing Task	22.9	21.34	22.5	As expected
Pan American	Performance Task	22.5	19.57	19.8	As expected
	Analytic Writing Task	22.9	20.54	21.5	As expected
Permian Basin	Performance Task	22.5	21.97	20.9	As expected
	Analytic Writing Task	22.9	22.22	22.0	As expected
San Antonio	Performance Task	22.5	21.59	22.0	As expected
	Analytic Writing Task	22.9	22.30	23.1	As expected

Note: Freshman level data were not available for U. T. Austin, U. T. Brownsville, and U. T. Tyler

Within Institution Freshmen Results

UT Arlington

At UTA, the expected Performance Task score was 22.64; the actual score was 22.5, which did not differ in a statistically significant way from the expected score. The actual score was consistent with what one would expect given the national norms and the composition of the UTA student body. The actual UTA analytic writing scale scores were nearly identical to their expected scores (22.81 v. 22.90).

UT Dallas

In the performance task scale of the test, the freshmen from UT Dallas outperform the national sample and their expected scores. Their performance task scale score was 27.1, while their analytic writing task score was (26.7) which are higher that what would be expected of these students (26.21 and 26.3) and higher than the national sample. The UT Dallas freshmen outscored the national sample schools by more than a standard deviation in both sections of the test.

UT El Paso

The freshmen students at this institution scored below the national sample on the performance task score. The national sample scored 22.5 and the UTEP sample scored 20.4. However, given their ACT scores, the freshmen scored as expected. The expectation was that UTEP students would score at 20.29; the actual score in this section of the test was 20.4. On the analytic writing task scale, the freshmen did as well as the national sample. Yet, the actual score was higher than the expected score for these students. However, there are no statistical differences in either set of scale scores. The UTEP students did not differ from the national sample in any significant way.

UT Pan American

The freshmen for this institution scored lower than the national sample as a whole in both the performance task and analytical writing task scale scores. At UTPA the score for the performance task scale was 19.8, which is significantly below the national sample's average score of 22.5. However, the expected score in the performance task scale, given the students' preparation, was similar to the actual score. On the other hand, while the students' score in the analytical writing task scale was lower than the national sample score of 22.5, the UTPA freshmen students scored better than expected. Yet, the difference between the expected and actual scores was not statistically different.

UT Permian Basin

The sample of freshmen students in this institution scored below the national sample scores in the performance task scale. The UTPB student average score for this section of the test was 20.9. The national sample score was 22.5. Yet, the expected score which is based on the students ACT scores, was 21.97. Their actual scale score was 20.9. The differences between the actual and expected scores do not differ significantly. On the other hand, the scores for the national sample, the expected score, and the actual score were similar at 22.9, 22.2, and 22.0 respectively. The students are performing as expected and similar ways when compared with the national sample.

UT San Antonio

The data for the sample at UTSA show that the freshmen students are performing at the same level as the national sample as well as how they are expected to perform, given the students' academic preparation. The average scale score for the national sample was 22.5; the expected average score is 21.59; and the actual score in the performance task was 22.0. These scores do not differ statistically in any significant way from each other. In other words, the freshmen students are doing as well as expected at UTSA and in relation to the national sample. Concerning freshmen performance on the analytic

writing scale, the freshmen sample scores at the same level as the national sample, the expected scores, and the actual scores.

2. UT System seniors from two institutions (UT Austin and UT Dallas) outperform the national sample.

In summary, Table 2 shows the comparison of senior CLA score ranges for UT System institutions with the national sample group. This table indicates that senior students at UT System institutions perform as well as or better than other institutions in the national sample. UT Austin and UT Dallas are performing better than the national group in absolute scores in both tests.

Regarding student growth on the analytical writing test, UT San Antonio, UT Pan American, UT El Paso, and UT Austin add significant value to student learning in this area. UT Arlington, UT Dallas, and UT Tyler did as well as other institutions around the nation.

Concerning the Performance Task test scores, UT San Antonio and UT Pan American are adding significant value to student development in this area. The rest of our institutions, UT Arlington, Austin, Dallas, and El Paso, perform within expected levels given the academic preparation of their students. Senior level data were not available for UT Permian Basin, UT Tyler, and UT Brownsville.

Table 2

University of Texas System
Senior-Level CLA Scores by Institution

Institution	Measure	National Sample Average Score	Average Expected Institution Score	Actual Average Institution Score	Actual Senior Performance Relative to Expected Performance	National Comparison
Arlington	Performance Task	24.8	24.74	24.3	-0.44	As expected
	Analytic Writing Task	27.3	27.08	27.2	-0.06	As expected
Austin	Performance Task	24.8	28.05	27.7	-0.35	As expected
	Analytic Writing Task	27.3	30.01	30.9	0.89	As expected
Dallas	Performance Task	24.8	28.59	29.0	0.41	As expected
	Analytic Writing Task	27.3	30.98	31.1	0.12	As expected
El Paso	Performance Task	24.8	23.76	23.0	-0.76	As expected
	Analytic Writing Task	27.3	25.90	27.3	1.40	Above expected
Pan American	Performance Task	24.8	22.70	23.5	0.80	As expected
	Analytic Writing Task	27.3	24.70	25.9	1.20	Above expected
San Antonio	Performance Task	24.8	23.69	25.0	1.31	Above expected
	Analytic Writing Task	27.3	27.06	28.3	1.24	Above expected
Tyler	Performance Task	NA	NA	NA	NA	Not available
	Analytic Writing Task	27.3	27.71	28.4	0.69	As expected

Note: Senior level data were not available for U. T. Brownsville and U. T. Permian Basin

Within-Institution Senior Analysis

The information that follows in the next table shows the statistics related to seniors presented in Table 2. Additionally, this section provides the institution's context concerning total enrollment, income, and student characteristics such as enrollment status, and the proportion of students receiving financial aid.

UT Arlington

The University of Texas at Arlington enrolls 25,297 students[1]. In fall 2004, there were 2,072 first-time undergraduate students[2]. The average SAT score of entering students in fall 2004 was 1072. Tarrant County, in which UT Arlington is located, has a median family income of $47,660. Twenty-eight percent of the student body is part-time[3], and 30% of the undergraduate student body received need-based financial aid in 2004[4].

UTA seniors did as well as the national sample of seniors in the performance task scale of the test. The national expected score was 24.74; while the senior actual test-score (24.3) was at the same level as the national score. Similarly the expected score and the actual score were at the same level. This means that senior students at UTA are doing as well as expected when compared with the national sample and their expected scores. Concerning the difference between the expected and the actual scores, there is a slight non significant difference. On the other hand, the senior scores in the analytical writing test are all similar among the national sample average score, the expected average score and the actual average test score.

UT Austin

The University of Texas at Austin enrolls 50,377 students. In fall 2004, there were 6,782 first-time undergraduate students. The average SAT score of entering students in fall 2004 was 1230. Travis County, in which UT Austin is located, has a median family income of $45,245. Nine percent of the student body is part-time, and 33% of the student body received need-based financial aid

The seniors outperformed the national sample in the performance task test. The national average score was 24.8; while the senior actual average test score was 27.7. Yet when one compares the expected score, which is based on students' academic preparation (ACT scores), against the actual score, the seniors underperformed slightly. Those differences, however, are not statistically significant. On the other hand, the seniors outperformed in a significant way the national group in the analytical writing test. The national sample average score was 27.3; while the seniors' actual performance in the writing test was 30.9. Moreover, the difference between the expected against the actual scores is slightly positive.

[1]The University of Texas System Accountability and Performance Report 2004-2005, Office of Institutional Planning and Accountability, http://www.utsystem.edu/IPA/acctrpt/2004/studentaccess.pdf

[2] Statistical Handbook 2005, Office of Institutional Studies and Policy Analysis, http://www.utsystem.edu/isp/StatHndbk/2005/Students.pdf

[3] Part-time status calculated by the Office of Academic Affairs based on data obtained for the Statistical Handbook 2005 prepared by the Office of Institutional Studies and Policy Analysis, http://www.utsystem.edu/isp/StatHndbk/2005/Students.pdf

[4] Data obtained from the Office of Institutional Studies and Policy Analysis

UT Brownsville/Texas Southmost College

The University of Texas at Brownsville/Texas Southmost College enrolls 11,546 students. Cameron County, in which UT Brownsville is located, has a median family income of $26,330. Fifty-two percent of the student body is part-time, and nearly 70% of the undergraduate student body received need-based financial aid in 2004[5].

UT Brownsville did not have enough data for this analysis.

UT Dallas

The University of Texas at Dallas enrolls 14,092 students. In fall 2004, there were 1,167 first-time undergraduate students. The average SAT score of entering students in fall 2004 was 1235. Dallas County, in which UT Dallas is located, has a median family income of $41,147. Thirty-four percent of the student body is part-time, and 34% of the undergraduate student body received need-based financial aid in 2004.

The seniors at this institution outperformed in a significant way the national sample scores in the performance task test. The national average score was 24.8; while the seniors at UT Dallas scores 29.0. This is a statistically significant difference between those two sample scores. Similarly, the seniors outperformed their expected score which was 28.59; while, the actual score was 29.0. This means that seniors did better than expected; this may be related to their growth in this area. The same is the case when one analyzes the analytic writing test scores. The UTD seniors outperformed the national sample score which was 27.3; while the UTD senior average score was 31.1. The seniors also performed as well as they were expected, given their academic preparation.

UT El Paso

The University of Texas at El Paso enrolls 18,918 students. In fall 2004, there were 4,060 first-time undergraduate students. The average SAT score of entering students in fall 2004 was 916. El Paso County, in which UT El Paso is located, has a median family income of $31,086. Thirty-one percent of the student body is part-time, and 51% of the undergraduate student body received need-based financial aid in 2004.

The seniors at UTEP scored a lower than the national sample on the performance task scale. The national sample scored 24.8; while the UTEP sample of seniors scored 23.0. Yet, the seniors scored as well when one compares the expected score against the actual score on the performance task scale. There is a slight variation between those two scores; yet, such variation is not statistically significant. On the other hand, UTEP seniors scored at the same level as the national sample on the analytic writing scale (27.3). When one analyzes the expected versus the actual scores, however, UTEP seniors performed above expected scores. The deviation score was greater than expected. That means that seniors at UTEP have achieved higher scores in writing than two-thirds of the seniors nationally.

UT Pan American

The University of Texas at Pan American enrolls 17,030 students. In fall 2004, there were 2,823 first-time undergraduate students. The average SAT score of entering

[5] Data on financial aid obtained for the institution's Institutional Compact, FY 2006-2007, http://www.utsystem.edu/IPA/compacts/2005/UTB-TSC06-07Compact.pdf

students in fall 2004 was 805. Hidalgo County, in which UT Pan American is located, has a median family income of $25,894. Twenty-eight percent of the student body is part-time, and 57% of the undergraduate student body received need-based financial aid in 2004.

The seniors at UTPA scored below the national sample on the performance task scale. The national average score was 24.8; while the score for UTPA is 23.5. The differences, however, are not significant in any way. However, there is a difference between the expected scores in this area and the actual score on the performance task scale. The expected score, which is based on the students ACT scores, is 22.70; while, the actual scale score is 23.5. That means that seniors have done better than expected on the performance task test. Similarly, on the analytic writing scale, the national group outperformed UTPA seniors. The national group scale score is 27.3; while UTPA's sample of students is 25.9. More important, the seniors outperformed their expected scores in this scale by a significant difference. The expected scale score is 24.7; while their actual score is 25.9. The seniors scored above expectations in the analytic writing test. That indicates that significant change has taken place in student learning in this area.

UT Permian Basin

The University of Texas of the Permian Basin enrolls 3,291 students. In fall 2004, there were 265 first-time undergraduate students. The average ACT (SAT) score of entering students in fall 2004 was 996. Ector County, in which UT Permian Basin is located, has a median family income of $33,045. Thirty-seven percent of the student body is part-time, and 43% of the undergraduate student body received need-based financial aid in 2004.

UT Permian Basin did not have enough seniors in the data sample to calculate their scores.

UT San Antonio

The University of Texas at San Antonio enrolls 26,175 students. In fall 2004, there were 4,421 first-time undergraduate students. The average SAT score of entering students in fall 2004 was 1006. Bexar County, in which UT San Antonio is located, has a median family income of $38,521. Twenty-five of the student body is part-time, and 38% of the undergraduate student body received need-based financial aid in 2004

The seniors at UTSA performed better than the national sample on the performance task scale. The national average score is 24.8 while the sample at UTSA scores 25.0. These are not significant differences; however, when one compares the expected score and the actual score for UTSA seniors, they outscored their expected performance by a significant portion. The expected score is 23.69 and their actual score is 25.0. This means that UTSA seniors added significant analytic writing skills to their knowledge while in UTSA. The same is the case when one analyzes their analytic writing scale scores. UTSA seniors outperformed their national peers 28.3 against 27.3. Moreover, UTSA seniors scored significantly better than their expected scores. UTSA seniors scored above expected in the writing skills test.

UT Tyler
The University of Texas at Tyler enrolls 5,326 students. In fall 2004, there were 521 first-time undergraduate students. The average SAT score of entering students in fall 2004 was 1068 [THECB data appear to be incorrect also, so I can't verify this statistic]. Smith County, in which UT Tyler is located, has a median family income of $38,561. 36% of the student body is part-time, and 40% of the undergraduate student body received need-based financial aid in 2004.

UT Tyler seniors only had enough data in the test of writing skills. The seniors outperformed the national sample, although the difference is not significant (28.4 versus 27.3). The seniors also outperformed their expected scores in the analytic writing test. That means that there is positive growth in student development in the writing achievement.

Summary

UT System academic institutions do as well or better than the national sample in terms of the how seniors and freshmen perform in the CLA performance task, which measures problem solving, critical thinking, and analytical reasoning. Seniors from UT San Antonio, Pan American, and Dallas do particularly well when compared with the national sample. On the other hand, when assessing the analytic writing task scores, seniors at El Paso, San Antonio, Pan American, Austin, Tyler, Dallas, and Arlington, do as well or better than the national sample. Finally, it is quite clear that Permian Basin, San Antonio, Pan American, and Arlington add significant value to their senior students when freshmen and senior score differences are taken into consideration.

How Will Test Results be Used

Chief academic officers may use the test results to address weaknesses in their general curriculum or to build opportunities to improve skills critical thinking, problem solving, analytical reasoning, and writing skills in the overall undergraduate preparation program. Chief academic officers may also use these test results for benchmarking academic performance of their students against national peers and setting targets for improvement.

Furthermore, chief academic officers may use these results to provide information to the public, funding organizations, policymakers, and parents on how their students perform academically in relationship to a national standard.

VI. Conclusions

With the advent of information technology, access to learning opportunities is greater now than ever. And postsecondary organizations are not the only ones providing such learning opportunities. In fact, other organizations have made significant inroads by providing performance-based learning opportunities. It is now possible for sophisticated consumers to obtain skills through different modes of instruction and different times for delivery. Therefore, university leaders must begun to develop programs that can articulate the knowledge, skills, and abilities students are expected to learn and the

competencies required for the application of learned curriculum. Thus, we recommend that institutions of higher education be exhorted to develop performance-based initiatives to measure student learning.

Performance-based initiatives are important to communicate to students which competencies are important for them to attain and the extent to which their learning experiences are meeting those expectations. These initiatives are also important to communicate employers or the general public what students know and are able to do. In higher education, we typically talk about knowledge, skills, abilities, and competencies as being one and the same. For example, we speak of competent mathematicians and knowledgeable mathematicians. Yet, skills and knowledge are acquired through learning experiences; the different combinations of skills and knowledge one has acquired in a given program define the competencies an individual possesses. These competencies are acquired through integrative learning experiences provided by academic programs. Finally, different competencies are combined to perform or carry out a task. To put it simply, competencies are complementary phenomena that combine skills, abilities, and knowledge.

Therefore, we recommend that at least four general intellectual skills often assumed to be improved by liberal arts programs be assessed in higher education. We recommend these skills be measured consistently across institutions to facilitate comparisons and fulfill accountability requirements.

1. Critical Thinking Skills
2. Problem Solving Skills
3. Analytic Reasoning Skills
4. Writing Communication Skills

References

ACT. (2005). *Assessments.* Available at http://www.act.org/edcareer/assess.html

Astin, A. W. (1977). *Four critical years: Effects of college on beliefs, values, and knowledge.* San Francisco: Jossey-Bass.Astin, A. W. (1987). Assessment , value-added, and educational excellence. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain*. *New Directions for Higher Education, 59*, 29-31. San Francisco, Jossey-Bass.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education.* New York: American Council on Education/Macmillan.

Banta, T.W. (2002). *Building a scholarship of assessment.* San Francisco: Jossey-Bass.

Brunner, R. D. (1997). Raising standards: A prototyping strategy for undergraduate education. *Policy Sciences, 30*, 167-189.

Business-Higher Education Forum. (2004). *Public accountability for student learning in higher education: Issues and options*. Washington, DC: American Council on Education.

Callan, P. M., & Finney, J. E. (2002). Assessing educational capital: An imperative for policy. *Change*, *34*, 25-31.

Erwin, T. D., & Wise, S. L. (2002). A scholar-practitioner model for assessment. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 67-81). San Francisco: Jossey-Bass

Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment.* San Francisco: Jossey-Bass.

Gray, P. J. (2002). The roots of assessment: tensions, solutions, and research directions. T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 49-66). San Francisco: Jossey-Bass.

Hanson, G. R. (1988). Critical issues in the assessment of value added in education. In T. W. Banta & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 53-67). San Francisco: Jossey-Bass

Kerschner, L. (1987). Outcomes assessment in the California master plan. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain*. *New Directions for Higher Education, 59,* 29-31. San Francisco, Jossey-Bass

Klein, S., Kuh, G., Chun, M., Hamilton, L., & Shavelson, R. (In press). An approach to measuring cognitive outcomes across higher-education institutions. *Research in Higher Education.*

Krueger, D. W., & Heisserer, M. L. (1987). Student outcomes assessment: What institutions stand to gain. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New Directions for Higher Education, 59,* 45-56. San Francisco: Jossey-BassKuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change, 35*(2), 24-32.

Lenning, O. T. (1988). Use of noncognitive measures in assessment. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils. New Directions for Institutional Research, 59,* 41-51. San Francisco: Jossey-Bass

Lingenfelter, P. E. (2003). Educational accountability: Setting standards, improving performance. *Change, 35*(2), 19-23.

Maki, P. L. (2002). Developing an assessment plan to learn about student learning. *Journal of Academic Librarianship, 28*(2), 8-13.

Palomba, C. A. (2002). Scholarly assessment of student learning in the major and general education. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 201-222). San Francisco: Jossey-Bass.

Pascarella, E. T. (2001). Identifying excellence in undergraduate education. *Change, 33,* 18-23.

Peterson, M. W., & Vaughan, D. S. (2002). Promoting academic improvement: Organizational and administrative dynamics that support student assessment. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 26-48). San Francisco: Jossey-Bass.

Pike, G. R. (2002). Measurement issues in outcomes assessment. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 26-48). San Francisco: Jossey-Bass.

Shavelson, R. J., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change, 35*(1), 10-19.

University of Texas System. (n.d.). *Student learning assessment virtual center.* Austin: Author. Available at http://www.utsystem.edu/aca/assessment

Washington Advisory Group. (2004). Home page. Available at http://www.theadvisorygroup.com/index2.html