Voluntary System of Accountability (VSA)

Information on Learning Outcomes Measures[1]

Introduction

An overarching goal of the VSA is to improve public understanding about the function and operations of public colleges and universities while affirming the significance of the great variety of approaches and missions of U.S. higher education. To that end, the VSA strives to lend transparency and consistency in the way in which universities report student learning and student development through a common web reporting template – the College Portrait. The VSA learning outcomes tools described here are intended to be employed as part of a comprehensive system of assessment that may include locally developed assessments of learning (e.g., portfolios), measures of engagement, direct measures of learning, student satisfaction and tests of discipline knowledge.

The VSA Learning Outcomes Work Group was charged with identifying a small number of educational outcome assessment instruments that have the capability of measuring improvement in student abilities in critical thinking, analytical reasoning and written communication. After an extensive review of sixteen candidate instruments, the Work Group recommended three instruments and selected components for inclusion in the VSA. These are:

- **Collegiate Assessment of Academic Proficiency (CAAP) –** two modules: critical thinking and writing essay http://www.act.org/caap/
- **Collegiate Learning Assessment (CLA) –** complete test including performance tasks, analytic writing tasks http://www.cae.org/content/pro_collegiate.htm
- **Measure of Academic Proficiency and Progress (MAPP) –** two sub scores of the test: critical thinking and written communication http://www.ets.org/

The instruments were selected for the VSA in part because of the willingness of the vendors to respond to the goals of the VSA, to consider the adoption of common statistical strategies, to modify outputs and to provide consultation to colleges and universities regarding the administration of the tests and the intended uses and limitations of the test.

Measuring learning outcomes as part of the VSA is designed as a pilot project. Best practices will emerge as more institutions use the various instruments and more data are collected. The widespread use across different types of institutions will help to establish the interrelationships between the test instruments.

In addition, a higher education consortium of AASCU, AAC&U, and NASULGC were awarded a $2.4 million FIPSE grant from the Department of Education to systematically study different approaches to student learning assessment. The project will examine the multiple purposes of learning assessment and test the validity, comparability, and appropriate uses of a variety of assessment approaches. It builds on the existing learning outcomes framework provided by AAC&U's Liberal Education and America's Promise (LEAP) initiative, the AASCU/NASULGC Voluntary System of Accountability (VSA) project, and AASCU's American Democracy Project.

[1] *Excerpted from working papers of the Learning Outcomes Technical Work Group and updated based upon recommendations from the Presidential Advisory Committee.*

COMMON VSA PROTOCOL

The procedures outlined below are minimum, general guidelines. Specific methodological decisions should be made in consultation with the testing organization based on the individual situation at each institution. The guidelines are not intended to replace advice from the testing organizations.

Skills/Abilities Tested

The learning outcomes measured as part of the VSA are general cognitive skills that cut across disciplines. The intent is not to measure everything but to measure learning that is common, multidisciplinary, and university-wide. Measuring learning outcomes at the institution level should complement, not replace, assessments that are grounded in the disciplines or focused on assessment of general education. Certain types of written communication, critical thinking, and analytic reasoning skills are discipline specific. Such skills also have more general components that are not directly tied to a particular course of study, but are still important outcomes of higher education. This latter set of skills is essential in a world where factual knowledge is becoming increasingly obsolete. The focus of the VSA is to assess how an institution is teaching students broad communication, thinking, and reasoning skills.

For the three instruments selected, the testing organizations agreed each of their test instruments measures critical thinking, analytic reasoning, written communication broadly defined. The MAPP and CAAP critical thinking modules necessarily encompass analytic reasoning. The MAPP and CAAP each have separate written communication modules. The CLA instrument is designed to holistically assess higher order thinking skills such as critical thinking, analytic reasoning, problem solving, and written communication.

Groups Tested

The VSA focuses on institutional level learning outcomes for students when they enter an institution and when they exit an institution as well as the change that occurs between those two points. One type of student is the "traditional" college student who enrolls and graduates from one institution. Definitions for these student cohorts are outlined below.

- **Traditional First-time, Full-time Freshman Fall Cohort**
  The group of students entering in the fall term will be used for tracking purposes. This group includes all students who enter an institution as full-time, first-time degree-seeking undergraduate students during the fall term of a given year. A first-time undergraduate student is a student attending any institution for the first time at the undergraduate level. This also includes students enrolled in the fall term who attended college for the first time in the prior summer term, and students who entered with advanced standing (college credits earned before graduation from high school). *(Source: IPEDS Glossary)*

- **Traditional Seniors**
  Students in the university's traditional freshman cohort who, by accepted institutional standards, are expected to complete a bachelor's degree within six months of when the test is administered each spring. Students who have already been awarded their degrees should not be included in the potential sample pool.

Another increasingly common type of student is the transfer student that starts at one institution and finishes at a different institution. Measuring value added for transfer students will be optional for VSA participating institutions. As the methodology for measuring value added for transfer

students is not well-established, participating institutions will need to work closely with the test vendors – particularly if the transfer students do not have entering ability data, such as ACT or SAT scores. Suggested definitions for transfer student groups are below.

- **Entering Transfer Students**
  "Typical" transfer students who transfer from another higher education institution (4-year institution or community college) with **30 to 70 semester equivalent credit hours**, inclusive.

- **Transfer Seniors**
  Students who transferred in between **30 to 70 semester equivalent credit hours**, from another higher education institution and who, by accepted institutional standards, are expected to complete a bachelor's degree within six months of when the test is administered each spring. Students who have already been awarded their degree should not be included in the potential sample pool.

Methodology for Calculating Scores
The method for calculating value-added within the VSA is based on the methodology currently employed by the Council for Aid to Education (CAE) for their CLA instrument. Both ETS and ACT have committed to using the same methodology to report scores as part of VSA. ETS and ACT are currently planning studies to work out the specifications for their particular instruments with a target completion date of fall 2008.

Each testing company will employ Ordinary Least Squares (OLS) regression models that use—at the institution level—mean CLA, CAAP, or MAPP scores as the dependent variable and mean SAT or ACT scores as the independent variable. If a participating institution does not collect SAT (ACT) scores, ACT (SAT) scores are converted to the SAT's (ACT's) scale of measurement using a standard cross-walk. One equation will be run using freshmen's test scores and another equation using seniors' test scores. Both equations will use the institution as the unit of analysis. These equations establish the typical relationship between an institution's mean test score and the average academic ability (as measured by SAT/ACT performance) of its incoming students who participate in the testing program. These relationships will be used to predict an institution's mean test score.

Results are described in two ways for the VSA: as the learning gains between the freshman and senior years (or the value-added component); and as the actual average test scores for freshmen and seniors.

Learning gains or value-added scores reflect the difference between the actual and expected scores of graduating and entering students, taking into account the academic ability of the students. Each of the three testing organizations will use the same method to compute and characterize their learning gains or value-added scores for VSA purposes.

- **Well Below Expected** (2 or more standard errors below the expected score)
- **Below Expected** (Between 1 and 2 standard errors below the expected score)
- **At Expected** (Between -1 and +1 standard errors from the expected score)
- **Above Expected** (Between 1 and 2 standard errors above the expected score)
- **Well Above Expected** (2 or more standard errors above the expected score)

The reporting of the actual average scores demonstrates whether the average score of the seniors is higher than the average score of the freshmen. Since the range of scores varies

across the three instruments, their results do not allow for direct comparisons between instruments. (For an example display of the results on the College Portrait see http://www.voluntarysystem.org/about_cp/CollegePortraitExample_110507.pdf)

Testing Methods

To calculate value-added scores, the VSA requires a cross-sectional methodology rather than a longitudinal method. The work group and task force made this decision because the cross sectional method is quicker, simpler, and less costly to implement, and there was no evidence at this time that that the results from one or the other method was more valid or reliable. (Note: The test developers will continue to collect data on any differences between the two strategies. If it becomes apparent that the longitudinal method is clearly superior to cross sectional method and well worth the additional resources, that decision will be revisited.)

Sampling Method

For the VSA, the unit of analysis is the institution rather than individual students. Therefore, at a minimum, participant in the VSA should select a sample that is representative of the undergraduate population as a whole. If an institution is interested in results for particular subgroups (e.g., males and females, residential and commuting students, student in different academic units) the sample must contain enough students from each of these groups to produce valid results.

Basic, minimum standards for the sampling of students are established as a random sample of freshmen and a random sample of seniors. For sample size, institutions should follow the guidelines that have been recommended by the test developers. In general, ACT recommends a minimum sample size of 200 for the CAAP. A sample above 200 is clearly acceptable but institutions should consult with the testing organization to verify the larger sample is representative of the underlying populations.

Frequency of Administration and Reporting of Learning Outcome Results

Since the measurement of student learning at the institutional level is not widespread, many institutions will need a period of time to find the best methods of administration and to determine how to use the test results to improve their educational programs before making the results of the outcomes tests public. For a period of four years, institutions may choose not to publicly report test results. After the four-year period is concluded, institutions will report and update the results at least once every three years.

Estimated Direct Costs

Costs to administer tests vary based on the number of students surveyed and services provided by the testing organization. For VSA purposes, the costs would be incurred every three years. Cost estimates given below follow the sample size recommendations of the testing organizations to report institutional level scores. The costs are for illustrative purposes only and institutions should contact the testing organizations for exact costs and administration options. The costs do not include any participation incentives distributed by an institution or oversampling to obtain scores for specific subgroups.

- CAAP (200 students each in fall and in spring): $10,200
- CLA (100 students each in fall and in spring): $6,500
- MAPP (200 students each in fall and in spring): $6,200

DESCRIPTIONS OF SELECTED INSTRUMENTS

Collegiate Assessment of Academic Proficiency (CAAP)

(Link to CAAP information specific to the VSA: http://www.act.org/caap/vsa/)

Test Overview

CAAP, an ACT product, is the oldest of the three tests recommended for VSA and has been utilized by colleges and universities for 21 years. The CAAP modules were originally designed to help institutions evaluate the outcomes of their general education. CAAP offers six independent test modules: Reading, Writing Skills, Writing Essay, Math, Science, and Critical Thinking. Scores for the five CAAP objective tests are scaled from 40 to 80, with a national mean around 60 and standard deviation around 5. The scale for the Writing Essay is 1 to 6.

For VSA purposes, two test modules have been selected: Writing Essay and Critical Thinking.

Writing Essay

The CAAP Writing Essay Test is focuses on the skills commonly taught in college-level writing courses and required in upper-division courses.

- Formulating an assertion about a given issue
- Supporting that assertion with evidence appropriate to the issue, position taken, and a given audience
- Organizing and connecting major ideas
- Expressing ideas in clear, effective language

The Writing Essay Test is designed to elicit responses that demonstrate a student's ability to perform these skills. Two 20-minute writing tasks are defined by a short prompt that identifies a specific hypothetical situation and audience. An examinee is instructed to take a position on the issue and to explain to the audience why the position taken is the better (or best) alternative. In order to more clearly define the audience and provide a focus for responses, each prompt specifies the basis upon which the audience will make its decision. Situations and audiences defined in the writing prompts are constructed so that the required background knowledge and experience are within the command of college sophomores.

ACT has developed a six-point, modified holistic scoring system. Each essay is read by two trained raters who independently score the essay on a scale from 1 to 6 (1 being the lowest score, 6 the highest). The scores from the two raters for each of the two essays (four scores) are averaged for the reported score, which ranges from 1 to 6 in increments of .5. The two raters' scores for each essay must either agree or be adjacent to be averaged. If the raters' scores differ by two or more points, a chief scorer adjudicates and determines the reported score.

Each score point reflects a student's ability to perform the skills identified above. Essays are evaluated according to how well a student formulates a clear assertion on the issue defined in the prompt, supports that assertion with reasons and evidence appropriate to the position taken and the specified concerns of the audience, and develops the argument in a coherent and logical manner. A student obtains lower scores for not taking a position on the specified issue, for not developing the argument, or for not expressing those ideas in clear, effective language. A student who does not respond to the prompt is assigned a "not ratable" indicator rather than a score on the 1 to 6 scale.

Critical Thinking
The CAAP Critical Thinking Test is a 32-item, 40-minute test that measures students' skills in clarifying, analyzing, evaluating, and extending arguments. An argument is defined as a sequence of statements that includes a claim that one of the statements, the conclusion, follows from the other statements. The Critical Thinking Test consists of four passages that are representative of the kinds of issues commonly encountered in a postsecondary curriculum.

A passage typically presents a series of sub-arguments in support of a more general conclusion or conclusions. Each passage presents one or more arguments using a variety of formats, including case studies, debates, dialogues, overlapping positions, statistical arguments, experimental results, or editorials. Each passage is accompanied by a set of multiple-choice test items. A total score is provided for the Critical Thinking Test; no sub scores are provided.

Additionally, ACT can generate Content Analysis Reports that allow institutions to compare the bottom 25%, middle 50%, and top 25% of their cohort with the national reference group in the areas of analysis of elements of an argument, evaluation of an argument, and extension of an argument.

Additional Reporting
The CAAP Standard Reporting Package, which includes Student Score Reports, is provided to the institution. ACT encourages institutions to share Student Score Reports with the students, but this decision is made at the discretion of the institution.

In addition, ACT publishes the CAAP User Norms Booklet every fall, which provides norms for each of the six test modules based on type of institution (two-year / four-year), year of students tested (freshman / sophomore / junior / senior), and institutional ownership (public / private). All CAAP "User Norms" are calculated on a three-year rolling basis.

ACT also can provide customized normative reports, but a minimum of 5 institutions must be included in the reference group. CAAP Linkage Reports indicate growth in terms of results being "greater than expected," "expected," and "less than expected."

Linkage Reports contain an analysis of performance for students who tested with the ACT Assessment or COMPASS upon entry to college and CAAP tests after their general education coursework has been completed. Because content specifications of some pairs of ACT Assessment or COMPASS tests and the CAAP tests are similar, it is possible to track performance. The local cohort is compared to all students nationally who have taken either the ACT Assessment or COMPASS and the CAAP tests, with the percentages of the national reference group adjusted to reflect the same ACT (or COMPASS) distribution of the local cohort.

Collegiate Learning Assessment (CLA)
(Link to CLA information specific to the VSA
http://www.cae.org/content/pdf/CLAVSA_Fact_Sheet.pdf)

A. Test Overview
The Collegiate Learning Assessment (CLA) is offered by the Council for Aid to Education (CAE) and is designed to assess an institution’s value added to key higher order skills: critical thinking, analytic reasoning, problem solving, and written communication. The CLA is intended to test in a

holistic manner (through writing and task performance) complex higher level skills and not the accumulation of knowledge from curricula.

The test includes two types of performance and analytic writing tasks which require open-ended responses. There are no objective (e.g., multiple choices) questions. There are two testing methodologies: cross-sectional and longitudinal. The cross-sectional methodology yields separate performance levels (well below expected, below expected, at expected, above expected and well above expected) for a freshman cohort, a senior cohort and freshman-to-senior (value added) performance.

The test is administered to a sample of freshmen and seniors and compared against similar samples obtained from other similar institutions. The CLA tests two kinds of performance and analytic writing:

Performance tasks
These tasks require students to use an integrated set of critical thinking, analytic reasoning, problem solving, and written communication skills to answer open-ended questions about a hypothetical, but authentic problem. A typical question might ask a student to identify and compare strengths and limitations of alternative hypotheses, points of view, and courses of action on a particular problem, by looking at a variety of documents and data.

Analytic Writing Tasks
These tasks require students to "make-an-argument," "critique-an-argument," and write analytically. 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. These writing 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.

The Measure of Academic Proficiency and Progress (MAPP)

(Link to MAPP information specific to the VSA
http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=ff3aaf5e44df4010VgnVCM10000022f95190RCRD&vgnextchannel=f98546f1674f4010VgnVCM10000022f95190RCRD)

Test Overview

MAPP, an ETS product, is described as an integrated test of general education skills. The full MAPP assessment provides skills sub scores for critical thinking, reading, writing and mathematics. It also provides context based sub scores for humanities, social sciences and natural sciences. For the purposes of the VSA, only written communication and critical thinking sub scores will be utilized.

Both norm and criterion referenced scores are provided. Norm referenced scores are used for baseline performance or for institutional comparison while criterion referenced scores are beneficial for determining strengths and weaknesses in curriculum. As a result, MAPP can be utilized to provide institutional reporting as required by the VSA and, if desired, expanded to provide information for curricular improvement. Both paper and web based forms are available. Scores can be received by the student and/or institution immediately with the web based version.

MAPP can be administered in two forms; the long form requires 120 minutes, the short form, 40 minutes. For VSA purposes the short form is suitable. The institution can receive summaries with either form. The short form does not provide individual scores as only one-third of the test is administered.