ACT Report
Voluntary System of Accountability Learning Gains Methodology

Calculating Postsecondary Institutions' Value-Added Performance Based on Mean ACT Composite, CAAP Critical Thinking, and CAAP Essay Scores

We have developed a method for using mean ACT and CAAP scores to measure the value that postsecondary institutions add to students' academic performance. This method mirrors the methodology used by the College Learning Assessment (CLA) for the same purpose.

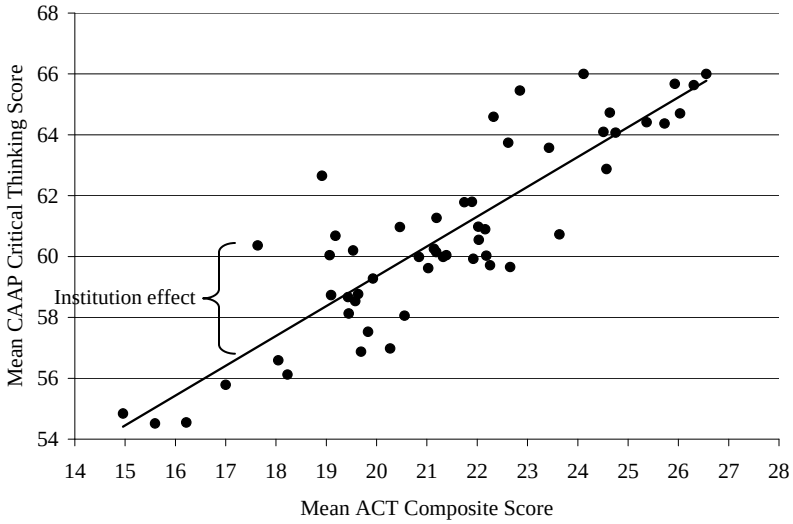
Step 1. Establish a relationship between ACT Composite Scores and CAAP Scores

The historical database of CAAP test scores was matched to the ACT database to find institutions where there were a sufficient number of jointly tested students. The results of this match are:

Test	Number of institutions
CAAP Critical Thinking (Freshman tested)	54
CAAP Critical Thinking (Senior tested)	76
CAAP Writing Essay (Freshman tested)	30
CAAP Writing Essay (Senior tested)	29

For each institution the mean ACT score and the mean CAAP score are calculated. These values are used to establish a regression equation. This equation is used to calculate an "expected" CAAP score for each institution. This is shown graphically in Figure 1 below. The graph indicates the relationship for freshmen testing with the Critical Thinking test.

Figure 1
Relationship of Mean ACT and Freshmen CAAP Critical Thinking Scores



The line represents the "expected" CAAP mean score and each institution's effect is the difference between the observed CAAP mean score and the expected CAAP mean score. This is shown in Figure 1 for a particular institution. The institution's ACT Composite average is just less than 18. Their expected CAAP mean score is 57, while their observed CAAP mean score is 60.2. The difference, 60.2 – 57 = 3.2 represents the effect of this institution for freshman. This will be referred to as the *freshman difference*. Using the same methodology described here for freshmen-tested students, we also derive the *senior difference*.

Step 2. Provide a measure of value-added

Provided that an individual institution tests an adequate number of freshmen and senior students with CAAP, the amount of value the institution added to students' academic performance can be computed as the *senior difference* minus the *freshmen difference*. This is the Value-Added score. There are five levels for this score.

a) *At expected*: Institutions whose value-added score is within one standard deviation of the mean (across all participating institutions) are said to have *at expected* performance.
b) *Above expected*: Institutions whose value-added score is greater than one standard deviation above the mean (but fewer than two standard deviations above the mean) are said to have *above expected* performance.
c) *Well above expected*: Institutions whose value-added score is greater than two standard deviations above the mean are said to have *well above expected* performance.
d) *Below expected*: Institutions whose value-added score is greater than one standard deviation below the mean (but fewer than two standard deviations below the mean) are said to have *below expected* performance.
e) *Well below expected*: Institutions whose value-added score is greater than two standard deviations below the mean are said to have *well below expected* performance.

An example of these calculations for a single institution is shown below.

Table 1. Example CAAP Value-Added Calculations

Example Data from an Institution – CAAP Critical Thinking	**Freshmen**	**Seniors**	**Value-Added**
Mean ACT score	20.46	20.50	
Expected mean CAAP Critical Thinking Score	59.80	62.11	
Actual mean CAAP Critical Thinking Score	60.97	62.40	
Difference (actual minus expected)	1.17	0.29	
Difference (actual minus expected), in standard deviation units	0.77	0.33	-0.43
Performance Level			***At expected***

Example Data from an Institution – CAAP Essay	**Freshmen**	**Seniors**	**Value-Added**
Mean ACT score	19.15	19.06	
Expected mean CAAP Essay Score	2.92	3.15	
Actual mean CAAP Essay Score	2.89	2.93	
Difference (actual minus expected)	-0.03	-0.22	
Difference (actual minus expected), in standard deviation units	-0.13	-0.91	-1.26
Performance Level			***Below expected***

Notes

Before assigning level of the value-added score, the score must be calculated in standard deviation units. To accomplish this, we take the difference between the standardized senior difference and the standardized freshmen difference. We then apply a final adjustment that puts the value-added score in standard-deviation units. This final adjustment depends on the correlation between the freshmen difference and the senior difference.

The regression used in Step 1 gives each institution identical weight in the model. This is somewhat misleading as the sample sizes are different for different institutions, and a larger sample will give a more reliable estimate of an institution's mean score. This problem is inherent in the institution-level model being used. This problem will be mitigated to some extent as we gather more data and the sample sizes are approximately the same (n=100 to 200) across institutions.

Sample size is also an issue in the value-added estimate: value-added scores are less precise for institutions when either of the sample sizes (freshman or senior) is small. One remedy to this problem would be to require minimum sample sizes for participating institutions.

The regression equations from which the expected scores are derived (Step 1) are based on all institutions for which there was sufficient data at either the freshman or senior level. Thus, the institutions used to establish the equation for freshmen and the equation for seniors are different. Theoretically, this should not be a problem, but we did notice some differences in model fit. As more institutions are added to the database, this should become less of an issue.

An Alternative Model

The institution-based model has some difficulties noted above. We are aware that the Council for Accountability in Education is experimenting with a student-level model. ACT believes this model is more appropriate. To check the difference, we ran both models on all institutions that provided both freshman and senior data. The results for the Critical Thinking test are shown in Table 2.

Table 2. Comparison of Institution Level and Student Level Analysis

Institution	Value-Added Estimate, Institution Level	Value-Added Estimate, Student Level	Assigned Level Institution Level	Assigned Level, Student Level	Standard Error, Student level
1	0.92	1.33	At Expected	Above Expected	0.57
2	-0.37	-0.41	At Expected	At Expected	0.84
3	-0.07	0.30	At Expected	At Expected	0.70
4	-0.65	-0.83	At Expected	At Expected	0.50
5	-0.25	-0.54	At Expected	At Expected	0.59
6	-2.71	-3.05	Well Below	Well Below	0.68
7	-1.63	-0.94	Below Expected	At Expected	0.82
8	1.53	2.02	Above Expected	Well Above	0.65
9	-0.55	-0.45	At Expected	At Expected	0.46
10	0.46	0.50	At Expected	At Expected	0.40
11	2.08	2.34	Well Above	Well Above	0.61
12	1.03	0.71	Above Expected	At Expected	0.64
13	-0.42	-0.60	At Expected	At Expected	0.49
14	-0.06	0.05	At Expected	At Expected	0.49
15	0.11	-0.97	At Expected	At Expected	0.41
16	0.19	0.62	At Expected	At Expected	0.30
17	-0.06	0.27	At Expected	At Expected	0.52
18	0.10	0.27	At Expected	At Expected	0.83
19	0.45	0.69	At Expected	At Expected	0.38
20	0.34	0.26	At Expected	At Expected	0.79

As can be seen in Table 2, the two estimated Value-Added scores for an institution are close to one another. The levels for each are usually the same, with the cases that are different highlighted in yellow. In each of these cases, it is an instance where the institution is very close to the border for a level on one method or the other. The different assignment of levels in those cases can be attributed to chance.

An important advantage of the student level model is that it provides an estimate of the standard error of the Value-Added score. This is shown in the last column. These values are larger than we might like. If there was an institution with an estimated Value-Added score of just above 1, and its standard error was more than .5, then we would classify this institution as Above Expected, even though the estimate is not significantly different from 0. Thus, we might see an institution swing from Above Expected to Below Expected in a year, without any real change. In order to make sure this doesn't happen, ACT would propose that we use this alternative model in order to ensure that the standard errors are small enough that the classification into levels will be consistent from year to year. Note that these standard error estimates are not available from the institution level models, as the error for the estimate of the Value-Added score cannot be separated from error in the model.

Recommendations

- Going forward, a plan will need to be put into place that emphasizes to the institutions that they will need to provide ACT with ACT test scores for all students that participate in VSA testing. SAT CR+M or SAT CR+M+W is also acceptable, but must be kept separate (from each other). We should not match names of those tested to our database of ACT test scores, as this will be too time consuming and expensive.
- Data from 13 additional colleges will be available in January 2009 (FIPSE TVS study). While the sample sizes are small, they will add to the base needed for creating stable regression equations. When this data is received, the models should be updated.
- A policy for dealing with schools with insufficient samples needs to be considered. The standard errors can be used as a method for determining the sample size needed. It would be our recommendation that the target be a standard error of .5 (less is preferable), and come up with a sample size recommendation based on what is likely to achieve the standard. If a school delivers less than that agreed upon number, then they would not get a classification. This policy would need to be communicated to the institutions, so that they are aware of the consequences. At some point, we would need to decide if an institution that just fails to meet the criterion would get a classification. Note that we would devise sample sizes that vary by institution size, so that smaller schools are not overburdened.