Time on Test, Student Motivation, and Performance on the Collegiate Learning Assessment: Implications for Institutional Accountability

Dr. Braden J. Hosch
Director of Institutional Research and Assessment
Central Connecticut State University, New Britain, CT
hoschbrj@ccsu.edu

****

Using results from the Collegiate Learning Assessment (CLA) administered at Central Connecticut State University, a public Carnegie master's-larger programs university in the Northeast, this study demonstrates time on spent on the test, student motivation, and to a lesser extent the local institutional administration procedures represent problematic intervening variables in the measurement of student learning. Findings from successive administrations of the instrument reveal wide year-to-year variations in student performance related to time on test and motivation. Significant additional study of these factors should likely be prioritized ahead of adoption of accountability practices that rely upon low-stakes testing to measure student learning and demonstrate institutional effectiveness.

****

Recent accountability initiatives in higher education have called for the direct assessment of student learning in ways that provide comparable information across institutions and states (Commission on the Future of Higher Education, 2006; Miller, 2006). Of particular note, the Voluntary System of Accountability (VSA) prompts public institutions to administer common standardized instruments to measure student learning and to examine value added by institutions to the educational experience (McPherson & Shulenburger, 2006). Such initiatives are laudable in that determining effectiveness of educational practices is essential to promote curricular and pedagogical improvements and to address changing learning styles and student populations. Current methods of assessing student learning using these instruments, however, may have uncertain value because they do not control for or acknowledge the fundamental issue of student motivation, especially in the context of a low-stakes test.

Using results from the Collegiate Learning Assessment (CLA) administered at Central Connecticut State University, a public Carnegie master's-larger programs university (Fall 2009 FTE enrollment, 80% undergraduate, 22% residential), this study demonstrates time on spent on the test, student motivation, and to a lesser extent the local institutional administration procedures represent problematic intervening variables in the measurement of student learning. Findings from successive administrations of the instrument reveal wide year-to-year variations in student performance related to levels of student motivation. Implications for accountability systems suggest that efforts should be directed to understanding what such test results mean more thoroughly before applying real principles of accountability, such as allocation of funds based on test result, to higher education institutions or systems.

Background

VSA requires participating institutions to administer one of three standardized instruments to measure student learning and to demonstrate the value-added to learning by the institution. These three instruments are the Collegiate Assessment of Academic Proficiency (CAAP) owned by ACT, Inc., the Measure of Academic Proficiency and Progress (MAPP) owned by the Educational Testing Service, and the CLA owned by the Council for Aid to Education.

The measurement construct for evaluating the value added by institutions adopts a cross-sectional design with institutions administering tests to samples of at least 100-200 first-year students and

100-200 graduating seniors who began their undergraduate experience at the institution. Scores on the tests are compared to an expected score based on SAT or ACT scores, and a relative-to-expected score is calculated as the residual between the actual and expected scores (performance categories are then described as "well above expected," "above expected," "at expected," "below expected," and "well below expected"). Further an institutional value-added score is calculated by subtracting the first-year residual from the senior residual (Klein, et al., 2007; Steedle, 2009). For instance, if entering first year students score at expected while seniors score well above expected the institution's value-added score will likely also be above or well above expected. Conversely, for institutions at which first-year students score above expected levels but seniors score at expected levels, the institutional value added might be below expected, depending on the magnitude of the score differential.

These methods have not been without research, critique, defense and unresolved controversy. CLA posts links to over a score of articles that describe and present the research behind the construct and methods of the assessment, including a recent validity study conducted by CLA and the other test owners that indicates the tests are valid and reliable (Klein, Liu, & Sconing, 2009). Still, CLA and the VSA have been criticized for use of a cross-sectional methodology to established educational value-added (Garcia, 2007; Banta & Pike, 2007; Kuh, 2006). Borden & Young (2008) provide an eloquent and comprehensive examination of the deployment of validity as a construct, using CLA and the VSA as a case study, to highlight the contextual and contested nature of validity across various communities. Testing organizations have tried to answer these charges (Klein, Benjamin, Shavelson, & Bolus, 2007), perhaps most effectively by demonstrating the utility of their instruments in longitudinal administrations to the same students (Arum & Roksa, 2008), although such practices can be prohibitively expensive and take years to produce results.

These debates are important and should continue, but they ignore the fundamental issue that these measurements are made with low-stakes tests and that variations in student motivation will have an effect on test scores. A wide set of studies has demonstrated that higher levels of motivation are associated with higher test scores for students at all levels. While there are some possible adjustments to control for this effect, such practices are speculative or require discrete item response analysis generally unavailable in constructed response assessments (Wise & DeMars, 2005; Wise S. L., 2006; Wise, Wise, & Bhola, 2006). Further, some studies have found no correlation between motivation and ability (not to be confused with test performance), suggesting such controls may be elusive (Perloff, 1964; DeMars, 1999). Despite the fairly commons-sense premise that students who apply little effort on a test may not perform well on it, educational accountability systems at all levels operate as though student motivation has little effect on test performance.

Caveat about Time on Test as a Proxy for Motivation

Much of this study considers time as a proxy for motivation, following the behaviorist premise that students spend time on activities as a function of their motivation to engage in the activity. For the student as *homo economicus*, time is a resource invested to maximize individual utility, be that for educational, recreational, economic, or other perceived benefit. Put more simply, students vote with their feet.

It should be acknowledged, however, that part of the skill of writing an answer to a timed essay test is to use time effectively to plan, organize, write, and revise an answer; students who may not have developed these skills or may not have worked through a complete and thorough solution to a problem will necessarily use less time to construct a response. Thus, it is important to observe

that time spent on the test includes a component of the skills tested for as well as a component of motivation for taking the test.

Methodology

For the present study, the CLA was administered to first-year students and seniors in 2007-08, 2008-09, and 2009-10 with the ultimate intention of publishing the scores on the institution's VSA College Portrait. Student recruitment has posed difficulties in all test administrations because many identified, eligible participants have balked at the prospect of taking a 90-minute essay test. Thus, while 683 students have been tested over three years, obtaining a representative sample of students to take CLA has constituted an ongoing challenge in test administration.

Recruitment of First-Year Students

Different strategies were employed to recruit first-year students and seniors. First-year students were recruited out of first-year experience (FYE) courses (maximum enrollment = 20), in which the faculty of record agreed to encourage students to participate in the CLA during a time outside of class, often scheduled individually in a computer lab or work station in the Office of Institutional Research and Assessment (OIRA).

Because of the political realities on campus of asking instructors to include CLA as part of their courses, the level at which it was integrated into the curriculum varied, as did the level of encouragement students received to participate. When CLA was highly integrated into the curriculum, some instructors required students to take the test, write a reflective journal entry about the experience, and then engage in classroom discussion about the role of standardized testing in the educational experience. In other instances, CLA was only loosely related to the class, and instructors awarded a few points of extra credit to students for their participation. Unsurprisingly, as stakes for taking the test increase, so did participation rates, but even among first-year experience section for which CLA was required, participation rates ranged from as high as 95% to as low as 55%. Because students are more or less randomly assigned to first-year experience courses, this sampling method generated a population of test takers who were roughly representative of the entering first-year class, though they by no means constituted a random sample. This practice of recruiting first-year students remained consistent over the three years of test administration.

Recruitment of Seniors

Recruitment of seniors posed even more significant challenges, and recruitment practices evolved of the course of the first test administration. Prior to the first test administration, the institution made the deliberate decision not to offer monetary or other incentives to take CLA because of some unease with the message sent by compensating students to take a test and also because of some uncertainty about the effectiveness of such practices. Thus, in the initial recruitment of seniors in Spring 2008, students received an email from the OIRA Director inviting them to participate, with emphasis on the benefits to students of seeing how their performance compared to seniors nation-wide as well as the benefit to the institution of gathering useful information about student learning. This strategy was completely unsuccessful, with no seniors taking CLA between January 29 and March 10, 2008, despite several follow-up communications.

To adjust before the testing window closed, seniors were offered a $25 discount on graduation regalia and three faculty members teaching senior level capstones in management, psychology, and social work had students in their classes take the test during a regularly scheduled class meeting. These strategies yielded 55 seniors through the class-based administrations and another 50 students who took CLA on their own. Subsequent invitations to seniors to participate in CLA offered a full waiver of regalia fees ($35 in 2009 and $40 in 2010) in the initial invitation and all

follow-up communications. Again, these procedures did not yield random samples, but in Spring 2009 and Spring 2010, students participating were roughly representative of the graduating class, with 41-45 majors represented in each term (compared to just 24 different majors in Spring 2008) and clusters of 10-12 students in expected areas in which students earn the highest portion of degrees (business, education, and psychology).

Proctoring of CLA in 2007-08 was conducted by a part-time temporary employee, who was a retired counselor from a local high school; some sessions were proctored by the OIRA Director. In 2008-09 and 2009-10 all sessions were proctored by a graduate assistant who had taken the CLA as a senior in Spring 2008; she had also lived on campus and served as president of an academic interest club and had numerous connections with undergraduate students. Beginning in 2008-09, students were asked to complete a nine-question survey prior to the test.

Table 1. Administration Details

Semester	Selection Method	Incentive	Proctor
Fall 2007	Students in FYE sections of instructors willing to participate	Course-based (required, extra credit, encouragement)	Retired HS Guidance Counselor / OIRA Director
Spring 2008	Email invitation and follow-up by OIRA Director; Three senior capstone sections (Management, Psychology, Social Work)	None (thru Mar. 10) $25 discount on graduation regalia (after Mar. 10) Capstone requirement (plus $25 discount above)	Retired HS Guidance Counselor OIRA Director
Fall 2008	Students in FYE sections of instructors willing to participate	Course-based (required, extra credit, encouragement)	Graduate Assistant who took CLA as senior in Spring 2008
Spring 2009	Email invitation and follow-up by OIRA Director; follow-up by graduate assistant	Waiver of entire graduation regalia fee ($35)	Graduate Assistant who took CLA as senior in Spring 2008
Fall 2009	Students in FYE sections of instructors willing to participate	Course-based (required, extra credit, encouragement)	Graduate Assistant who took CLA as senior in Spring 2008
Spring 2010	Email invitation and follow-up by OIRA Director; follow-up by graduate assistant	Waiver of entire graduation regalia fee ($40)	Graduate Assistant who took CLA as senior in Spring 2008

The amount of time students spent on the CLA was not collected until midway through the Spring 2008 administration following the observation that some students finished the exercise and left the room after a short amount of time (the shortest recorded time was 11 minutes) on activities that range between 60 to 90 minutes maximum. Through the Spring 2009 administration, CAE did not provide data about how much time participants spent on the assessment, and so the proctor kept track of the number of minutes students spent between the time the assessment was activated and the time they finished. This hand-recorded time includes time spent on the pre-test survey and the tutorial, and so has some amount of error in it.[1]

[1] Beginning with the Fall 2009 data received in March 2010, CAE began to include the time spent on CLA. The difference between locally kept time and computer-recorded time averaged about 7 minutes for Fall 2009. The exceptions were students with scaled scores over 1200 or with relative-to-expected scores in the well above expected

Findings

Results varied among administrations even though differences in testing populations were minor. The scaled score of first-year students increased from 1057 (51st unadjusted percentile, 62nd adjusted percentile) in 2007-08 to 1127 (67th unadjusted percentile, 84th adjusted percentile) in 2008-09 and then dropped to 1098 (53rd unadjusted percentile, adjusted percentile not available as of the presentation of this paper). Similarly the scaled scores of seniors increased from 1133 (37th unadjusted percentile, 63rd adjusted percentile) in 2007-08 to 1248 (70th unadjusted percentile, 98th adjusted percentile) in 2008-09; results from the Spring 2010 administration are unavailable, as of the presentation of this paper.

Table 2. Year-to-year participant profile and CLA performance

	2007-08	2008-09	2009-10
First-Year Students			
Students taking CLA (N)	105	110	130
HS rank percentile (mean)	57	61	62
SAT score (mean)[2]	1019	1045	1019
Scaled CLA score (mean)[3]	1057	1127	1098
Unadjusted CLA percentile	51	67	53
Adjusted CLA percentile	62	84	--
Performance relative to expected score	At expected	Above expected	--
Minutes spent on CLA (mean)	--	49	44
End of semester GPA (mean)	2.73	2.86	2.87
Cum GPA at end of semester (mean)	2.73	2.87	2.88
Seniors			
Students taking CLA (N)	99	134	105
HS rank percentile (mean)	64	63	
SAT score (mean)	994	1016	1045
Scaled CLA score (mean)	1133	1248	
Unadjusted CLA percentile	37	70	--
Adjusted CLA percentile	63	98	--
Performance relative to expected score	At expected	Well above expected	--
Minutes spent on CLA (mean)	45	63	55
End of semester GPA (mean)	3.19	3.30	--
Cum GPA at end of semester (mean)	3.13	3.24	--
Institutional "Value Added"			
Adjusted percentile	49	79	--
Performance relative to other institutions	At expected	Above expected	--

range; locally-kept times averaged 9 to 11 minutes longer time than the computer-tracked time, suggesting they may have spent more time on the pre-test tutorial and survey than other students.

[2] CLA actually represents this figure as Entering Academic Ability (EAA), which in general is the combined math and critical reading SAT score. For students without these scores, the Scholastic Level Exam is substituted.

[3] Each student takes either a performance task or a writing task, and CLA reports these scores separately. They are combined here as a scaled score; correlations and regressions both produce about the same results when conducting statistical tests separately on each score, but the degrees of freedom are obviously reduced by half.

Table 3. Minutes Spent on CLA by Scaled Score and Term

			CLA Scaled Score			
			<= 1000	1001 - 1100	1101 - 1200	1201+
Fall 2007	Participants	N=104	37%	22%	29%	13%
First-Year	**Minutes on CLA**	**Mean**	**NA**	**NA**	**NA**	**NA**
		St. Dev.	NA	NA	NA	NA
Spring 2008	Participants	N=99	20%	17%	30%	33%
Seniors	**Minutes on CLA**	**Mean**	**38.3**	**40.6**	**46.1**	**52.2**
		St. Dev.	19.1	20.4	13.3	11.2
Fall 2008	Participants	N=110	15%	23%	35%	26%
First-Year	**Minutes on CLA**	**Mean**	**36**	**47.3**	**51.1**	**54.4**
		St. Dev.	14.9	12.7	15.5	13.5
Spring 2009	Participants	N=134	6%	8%	28%	58%
Seniors	**Minutes on CLA**	**Mean**	**53.5**	**60.5**	**59.6**	**65.1**
		St. Dev.	20.9	25	22.5	18.5
Fall 2009	Participants	N=127	27%	22%	31%	20%
First-Year	**Minutes on CLA**	**Mean**	**32.6**	**40.1**	**49.8**	**56.6**
		St. Dev.	12.9	12.8	19	15.3
All First-Year	Participants	N=342	26%	22%	32%	20%
	Minutes on CLA	**Mean**	**33.7**	**43.5**	**50.5**	**55.4**
		St. Dev.	13.6	13.2	17.2	14.3
All Seniors	Participants	N=232	12%	12%	28%	47%
	Minutes on CLA	**Mean**	**43.8**	**52.1**	**56.5**	**63.4**
		St. Dev.	20.7	24.7	21.4	18.2

Table 4. Minutes spent on CLA by Relative-to-Expected Score and Term

			CLA Relative-to-Expected Score		
			Well Below and Below Expected	At Expected	Well Above and Above Expected
Fall 2007	Participants	N=104	32%	21%	47%
First-Year	**Minutes on CLA**	**Mean**	**NA**	**NA**	**NA**
		St. Dev.	NA	NA	NA
Spring 2008	Participants	N=99	32%	26%	42%
Seniors	**Minutes on CLA**	**Mean**	**34.8**	**47.8**	**50.8**
		St. Dev.	19.3	13.4	12.4
Fall 2008	Participants	N=110	24%	18%	58%
First-Year	**Minutes on CLA**	**Mean**	**41.9**	**49.2**	**51.5**
		St. Dev.	16.7	14.1	14.5
Spring 2009	Participants	N=134	21%	16%	63%
Seniors	**Minutes on CLA**	**Mean**	**57.5**	**58.3**	**65.0**
		St. Dev.	19.2	20.9	20.5
Fall 2009	Participants	N=127	34%	16%	50%
First-Year	**Minutes on CLA**	**Mean**	**33.1**	**47.9**	**51.0**
		St. Dev.	13.4	16.1	17.2
All First-Year	Participants	N=342	30%	18%	52%
	Minutes on CLA	**Mean**	**36.4**	**48.5**	**51.3**
		St. Dev.	15.3	15	15.8
All Seniors	Participants	N=232	26%	20%	54%
	Minutes on CLA	**Mean**	**49.2**	**54.3**	**62.8**
		St. Dev.	22	18.9	20.1

While differences among administration methods undoubtedly had some effect on results, the amount of time that students spent taking the test was strongly related to test performance. First-year students who achieved a scaled CLA score over 1200 spent an average of 55.4 minutes taking the test, while those who earned a scaled score below 1000 spent an average of 33.7 minutes on the test. Among seniors, those who achieved a scaled score over 1200 spent an average of 63.4 minutes on the test, while those who earned a score below 1000 spent an average of 43.8 minutes on the test.

When controlling for academic inputs by comparing actual CLA scores to expected CLA scores, a similar pattern emerges; students who spent more time on the test outperformed their expected score (based upon combined SAT score).

Correlations

Correlations among scaled CLA score and other relevant items yielded different sets of related items, but minutes spent taking the test was significant ($p<0.001$) for both first-year students ($R=0.468$) and for seniors ($R=0.331$). In fact, for first-year students, the amount of time spent taking CLA exceed SAT scores in the correlation with CLA scaled scores, accounting for just under 22% of variance. Conversely, for seniors SAT scores exceeded time spent on test in the correlations with CLA scores. For seniors, the time spent on CLA accounted for just 9% of variance.

Table 5. Correlations between Scaled CLA Scores and Related Factors (First-Year Students)

	CLA Scaled Score	Minutes Spent on CLA	SAT Critical Reading	SAT Writing	SAT (Math + CR)	High School Rank
Minutes spent on CLA	.468	--	--	--	--	--
SAT Critical Reading	.333	.178	--	--	--	--
SAT Writing	.311	.162	.632	--	--	--
SAT (Math + CR)	.326	.186	.807	.610	--	--
High School Rank	.264	.227	.133	.161	.180	--
SAT Math	.201	.127	.331	.370	.824	.145

Table 6. Correlations between Scaled CLA Scores and Related Factors (Seniors)

	CLA Scaled Score	SAT (Math + CR)	SAT Math	SAT Critical Reading	End of term cum GPA	High School Rank
SAT (Math + CR)	.505	--	--	--	--	--
SAT Math	.479	.886	--	--	--	--
SAT Critical Reading	.409	.890	.576	--	--	--
End of term cum GPA	.400	.271	.285	.263	--	--
High School Rank	.338	.214	.178	.147	.336	--
Minutes spent on CLA	.331	.090	.210	.095	.177	.214

All correlations in Tables 6 and 7 are two-tailed (Pearson's R) and are significant at a level of at least $p<0.01$.

Regression Models

Only two variables yielded statistically significant results in regression models: SAT scores and the number of minutes spent on the test. The number of minutes spent on CLA improved model power by about 8% for seniors (from $R^2=0.236$ with SAT scores along to $R^2=0.321$ when including the number of minutes spent on the test) and by just over 20% for first year students (from $R^2=0.104$ with SAT scores alone to $R^2=.0.261$ when including minutes spent on the test).

Table 7. Multivariate Models of CLA Scores (First-Year Students)

First-Year Student CLA Scaled Score (Adj. R^2=0.261)	β	Std. Err.	t	Sig.
(Constant)	652	73.3	8.89	***
Minutes spent on CLA	3.67	.490	7.49	***
Combined SAT Score	0.281	.071	3.93	***

First-Year Student CLA Percentile (Adj. R^2=0.286)	β	Std. Err.	t	Sig.
(Constant)	-28.9	12.6	-2.28	***
Minutes spent on CLA	.685	.084	8.18	***
Combined SAT Score	.049	.012	3.94	***

*** Sig. at p<0.000; df=234.

Table 8. Multivariate Models of CLA Scores (Seniors)

Senior CLA Scaled Score (Adj. R^2=0.321)	β	Std. Err.	t	Sig.
(Constant)	395	93.9	4.20	***
Minutes spent on CLA	2.50	.535	4.67	***
Combined SAT Score	0.665	.091	7.34	***

Senior Student CLA Percentile (Adj. R^2=0.315)	β	Std. Err.	t	Sig.
(Constant)	-69.0	14.2	-4.85	***
Minutes spent on CLA	.372	.081	4.59	***
Combined SAT Score	.099	.014	7.19	***

*** Sig. at p<0.000; df=174.

Self-Reported Motivation

CLA allows institutions to add nine local questions prior to administration of the test. This number of items is not sufficient to replicate items from validated instruments to measure motivation, such as the Motivated Strategies for Learning Questionnaire (Pintrich & DeGroot, 1990) or the Academic Motivation Scale (Vallerand, et al., 1992). Further, while in general the survey results reveal interesting opinions and preferences among the testing population, they were not useful in elucidating patterns of test performance (see appendix for full results).

Participants were asked to respond to nine items, including the statement "I feel highly motivated to participate in this activity today," on a five-point Likert-type scale. Interestingly, just 34% of first year students agreed or strongly agreed with this statement (differences between Fall 2008 and Fall 2009 were negligible), while 70% of seniors agreed or strongly agreed with this statement. Nevertheless, observed differences in neither scaled scores nor time spent on the test rose to the level of statistical significance.

Other items asked participants about test anxiety, preference for essay test over multiple-choice tests, individual responsibility vs. institutional responsibility for student learning, and the utility of college rankings publications in selecting a college. While the times and test scores associated with these items were suggestive, the times and test scores associated with almost all of them did not exhibit statistically significant results on an ANOVA test. The only item that yielded statistically significant results was "students are responsible for learning material assigned by their professors"; both time and test scores were lower for students who disagreed with this item ($p<0.05$).

Implications

Findings from this study that the amount of time students spend on tests like the CLA has a significant impact on performance should inform accountability systems in higher education and perhaps other bands of the educational spectrum. The overarching points that students get out of their educational experiences what they put into them and that the amount of time and effort they apply to educational activities matter are hardly surprising, but these points are frequently overlooked when fashioning measures that purport to measure institutional effectiveness. Further research at institutions and also at testing organizations should be conducted to determine more broadly the scope of the motivation and time effects on student performance.

Sensitivity to Recruitment Practices and Testing Conditions

CLA test creators maintain that the instrument is "sensitive to the effects of instruction" (Klein, Shavelson, & Benjamin, 2007), but this study suggests that CLA (and likely other instruments) may exhibit sensitivity to recruitment practices and testing conditions. This observation is far from surprising, but it is important to recognize that the conditions under which students are recruited and tested vary across institutions. The extent to which these difference may affect scores presents opportunities to misinterpret test results as well as possibilities that institutions may have incentives to focus efforts and resources on optimizing testing conditions for a small few rather than improving learning for the many. More effort and support should be focused on standardizing selection, recruitment, and incentive strategies.

Longitudinal Testing

Longitudinal testing of the same students at different points in their educational careers may have some potential to control for motivational differences among students that are present in the cross-sectional design. The Social Science Research Council study (Arum & Roska, 2008) makes a more compelling case for validity and reliability when testing the same students on the CLA. This method is also recommended by Garcia (2007). Costs of such testing are likely prohibitive for many institutions, especially those with mobile student populations. Nevertheless, VSA could benefit from some consideration of how to substitute longitudinal testing practices for institutions that choose to pursue this route.

Multi-Year Moving Averages

Higher education institutions are rarely characterized as moving quickly or making rapid changes, and so sharp improvements or declines in scores on cross-sectional samples should in general not be observed. VSA currently requires institutions to update scores once every third year (providing some incentive to pick the most favorable score), but a more reliable portrayal might be gained from presentation of a multi-year moving average of student performance. Some consideration of adjustments to VSA reporting practices to allow reporting of multi-year averages seems warranted.

Statistical Adjustments

Statistical adjustments to scores should be explored. Much of this work (Wise & DeMars, 2005; Wise, 2006; and Wise, Wise, & Bhola, 2006) has focused on item elimination on multiple-choices tests based on time spent on an item. Applying such methods may be difficult if not impossible on a holistically scored constructed response instrument. Further, elimination of items where students spent little time or did not try hard damages the political utility of publishing the scores—institutions hardly benefit when a close reading of the methodology reveals students who performed poorly were essentially removed from the testing population.

Portfolios, Projects, Theses, and External Examiners

A prominent argument against using portfolios or course-based work products to evaluate student learning has been that the amount of time student spend preparing the portfolio represents an uncontrolled intervening variable, but findings from the current study suggest similar problems arise from testing; the timeframes are simply shorter. Portfolios, projects, theses, or even exams can be designed to illustrate students learning, but the real control should be external evaluation. The advent of electronic portfolios, the ease of file sharing, and very inexpensive electronic storage might suggest that the formation of consortia of institutions (not unlike athletic conferences) to evaluate student work pooled from member institutions could promote confidence in results and provide a meaningful examination of student performance in context.

Effects in Other Educational Systems

The effects observed in this study that time spent on the test affects performance is likely not limited to the higher education sector; in fact, it seems reasonable that test performance among students in elementary and secondary schools is also influenced by motivation and the stakes associated with the test (even in high-stakes testing, the incentives are structured to prompt students to try just hard enough not to fail). Such factors should be researched in considerably more depth, given the extent to which federal and state funding and policy-making has become linked to student test scores. To the extent that low test scores might represent non-performance in a cognitive domain, curricular adjustments can and should be made, but to the extent that low performance or performance gaps may reflect non-performance in a behavioral domain, substantively different adjustments would need to be made to improve performance.

New Britain, Connecticut
May *2010*

Works Cited

Arum, R., & Roksa, J. (2008). *Learning to reason and communicate in college: Initial report of findings from the CLA longitudinal study*. New York: Social Science Research Council.

Banta, T., & Pike, G. (2007). Revisiting the blind alley of value-added. *Assessment Update , 19* (1), 1-2.

Borden, M. H. & Young, J. W. (2008). Measurement Validity and Accountability for Student Learning. *New Directions for Institutional Research, Assessment Supplement 2007*.

Commission on the Future of Higher Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: U.S. Department of Education.

Council on Independent Colleges. (2008). *Evidence of learning: Applying the Collegiate Learning Assessment to improve teaching and learning in the liberal arts*. Washington, DC: Council on Independent Colleges.

DeMars, C. (1999, April 20-22). *Does the relationship between motivation and performance differ with ability?* Paper presented at the meeting of the National Council on Measurement in Education, Montreal, Quebec.

Garcia, P. (2007, November 14-16). *How to assess expected value-added: The CLA method*. Paper presented at the meeting of the California Association for Institutional Research, Monterey, CA.

Klein, S., Shavelson, R, Benjamin, R. (2007, February 8). Setting the record straight. *Inside Higher Ed*. Retrieved May 27, 2010 from http://www.insidehighered.com/views/2007/02/08/benjamin

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). Collegiate Learning Assessment: Facts and Fantasies. *Evaluation Review 31*, (5), 415-39.

Klein, S., Liu, O. L., & Sconing, J. (2009). Test Validity Study (TVS) Report. Voluntary System of Accountability. Washington, DC: Association of Public Land-Grant Universities.

Kuh, G. (2006). *Director's message in: Engaged learning: fostering success for all students*. Bloomington, IN: National Survey of Student Engagement.

McPherson, P., & Shulenburger, D. (2006). *Toward a Voluntary System of Accountability (VSA) for public universities and colleges*. Washington, DC: National Association of State Land Grants Universities and Colleges.

Merrow, J. (2006). My college education: Looking at the whole elephant. *Change, 38* (3), 8-15.

Miller, M. (2006). *Assessing college level learning*. San Jose, CA: National Center for Public Policy and Higher Education.

Perloff, R. (1964). The hypothesized contribution of test motivation variance to test validity. *Educational and Psychological Measurement, 24* (4), 783-88.

Pintrich, R. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33-40.

Steedle, J. (2009). Advancing institutional value-added score estimation. CLA White Paper. New York: Council for Aid to Education. Retrieved May 20, 2010, from http://www.collegiatelearningassessment.org/files/advancinginstlvalueadded.pdf.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C, & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in Education. *Educational and Psychological Measurement, 52* (4), 1003-1017.

Wise, S. L. (2006). An investigation of the differential effort received on a low-stakes computer-based test. *Applied Measurement in Education, 19* (2), 95-114.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10* (1), 1-17.

Wise, V. L., Wise, S. L., & Bhola, D. S. (2006). The generalizability of motivation filtering in improving test score validity. *Educational Assessment , 11* (1), 65-83.

Appendix: Survey Results with Time on Test and Test Scores (F '08, Sp. '09 & F'09)

		First-Year (Fall 2008 and Fall 2009)				Seniors (Spring 2008 only)			
CLA Survey Item		N	Pct	Avg Minutes	Avg CLA Score	N	Pct	Avg Minutes	Avg CLA Score
I feel highly motivated to participate in this activity today	Strongly disagree	10	4%	40	1071	1	1%	^	^
	Disagree	19	8%	45	1103	1	1%	^	^
	Neutral	123	53%	46	1109	37	28%	59	1240
	Agree	64	28%	48	1125	70	53%	63	1250
	Strongly agree	14	6%	51	1169	22	17%	67	1244
I perform better on essay tests than on multiple choice tests	Strongly disagree	35	15%	48	1089	9	7%	57	1220
	Disagree	73	32%	45	1102	24	18%	61	1222
	Neutral	74	32%	48	1127	51	39%	62	1292
	Agree	34	15%	46	1129	36	27%	66	1207
	Strongly agree	15	6%	45	1155	11	8%	65	1265
I prefer to take a test rather than write a paper	Strongly disagree	11	5%	42	1083	9	7%	67	1205
	Disagree	30	13%	46	1143	16	12%	62	1248
	Neutral	51	22%	45	1122	33	25%	62	1290
	Agree	81	36%	48	1103	45	34%	67	1233
	Strongly agree	54	24%	48	1115	28	21%	57	1239
I get so nervous when I take tests that I don't usually perform my best work	Strongly disagree	21	9%	48	1157	13	10%	60	1286
	Disagree	54	24%	50	1126	46	35%	67	1271
	Neutral	76	33%	48	1114	39	30%	62	1209
	Agree	53	23%	42	1092	31	24%	61	1261
	Strongly agree	24	11%	45	1112	2	2%	^	^
Students are responsible for learning material assigned by their professors	Strongly disagree	0	0%	^	^	0	0%	^	^
	Disagree	1	0%	^	^	2	2%	^	^
	Neutral	19	8%	37*	1022*	11	8%	54	1187
	Agree	106	47%	45*	1124*	61	47%	65	1271
	Strongly agree	100	44%	51*	1121*	57	44%	63	1233
Colleges and universities are responsible if students don't learn what they need to be successful after they graduate	Strongly disagree	14	6%	48	1168	5	4%	74	1235
	Disagree	66	29%	46	1097	36	27%	60	1233
	Neutral	98	43%	48	1114	40	31%	63	1271
	Agree	40	18%	47	1122	36	27%	64	1242
	Strongly agree	8	4%	43	1118	14	11%	62	1245
All college students should be required to pass a standardized exit test in order to graduate	Strongly disagree	45	20%	45	1109	17	13%	62	1254
	Disagree	86	38%	48	1137	63	48%	63	1280
	Neutral	61	27%	47	1093	30	23%	61	1226
	Agree	31	14%	46	1119	13	10%	69	1176
	Strongly agree	5	2%	42	1009	8	6%	61	1197
Students should use published college rankings (like US News and World Report) when deciding which school to attend	Strongly disagree	7	3%	41	1199	4	3%	69	1239
	Disagree	37	16%	47	1135	33	25%	61	1270
	Neutral	118	52%	46	1114	47	36%	62	1231
	Agree	62	27%	48	1096	42	32%	63	1260
	Strongly agree	3	1%	45	965	5	4%	70	1181
I plan to complete my bachelor's degree at CCSU	Strongly disagree	4	2%	41	1245	0	0%	^	^
	Disagree	8	4%	49	1152	0	0%	^	^
	Neutral	48	21%	45	1095	2	2%	33	1263

	Agree	82	36%	47	1109	20	15%	66	1282
	Strongly agree	86	38%	47	1118	109	83%	63	1242

* si.g at p<0.05 on ANOVA test. ^ n is too small to provide even suggestive results.