



Office of Institutional Research

To: College-Wide Assessment Team
From: Darby Hiller
Date: July 10, 2003
Subject: Spring 2003 CAAP After Action Report

The Collegiate Assessment of Academic Proficiency (CAAP) Writing Essay test was administered successfully to 214 students in March 2003. Overall, the NMC mean score was 3.4 on a 6-point scale (further defined below), with about 91% of NMC students scoring at or above the national average for two-year higher educational institutions.

Test Background and Scoring.

The Essay test is intended to measure students' ability in "formulating an assertion about a given issue; supporting that assertion with appropriate evidence; organizing and connecting major ideas; and communicating using good writing skills in such areas as mechanics, sentence structure, and command of the language." (CAAP Planning and Forms, p.3)

The scoring system for the essay test, developed by ACT, is a modified-holistic system "that ranges from 1 to 6 in increments of 0.5. Each score point reflects a student's ability to perform the writing task defined in a 'prompt'...Responses are evaluated according to how well the student (1) formulates a clear assertion on the issues raised in the prompt, (2) supports that assertion with reasons and evidence appropriate to the position taken and the specified concerns of the audience, and (3) develops the argument in a coherent and logical manner" with organization and connections between major ideas, and (4) expresses those ideas in clear, effective language. (CAAP User's Guide, p. 16)

The Essay test consists of two different prompts. Students have 40 minutes total to write a coherent essay for each prompt (20 minutes a piece). The prompt identifies a specific hypothetical situation and audience and poses an issue on which a student is to take a position. The essays were scored by ACT, which gives NMC the advantage of having external evaluators in our assessment process. The ACT rubric is outlined in the Appendix.

Demographics.

The student demographics of the NMC sample did not vary much from the student body as a whole. Whites made up 83.1% of the sample, with about 10% choosing not to respond to the ethnic background question. Nearly 60% of the sample was women. About 51% was between the ages of 16 and 20. Seventy-seven percent of the test takers reported that they were full time students. The freshman made up nearly 40% of the sample, and there were just about as many sophomores (34.6%). A majority of students believed their cumulative grade point average was a 3.0 or higher. (See supporting data tables below.)

The most frequently reported college majors are as follows: business (11.3%), education (19.2%), health professions (6.5%), social sciences (10.3%) and home economics (11.2%). Please take care in interpreting these percentages for college majors because the high number of responses may be related to the course in which the tests were administered (RSM101, PSY250, PLS101, HST112, GEO105, SOC101, ENG111, CUL211 CIS210, BIO220, and ACC121).

Results.

The NMC institutional mean score was 3.4, which is above the national mean of 3.0 but within one standard deviation (0.6). Exactly 35.2% of NMC students scored at the national mean, with 55.4% actually scoring above the national mean.

When broken down by demographic, women (mean 3.44) scored slightly higher than men (mean 3.25). Expectedly, the students self-reporting the highest grade point average (3.51 or above) performed the best (mean score 3.68), while those reporting their grade point average was between a 2.0 and 2.5 performed the worst (mean score 3.0) but still at the national average. Those selecting Communications as their college major scored the highest (mean 4.25) of all the majors. Biological Sciences students scored the second highest (mean 3.88). Those reporting that Office Management was their college major scored the lowest (mean 3.0) but still at the national average. The number of respondents in each college major is really too low to make any statistically significant generalizations.

When broken down by educational level (student reported credit hours), the students performed as expected. Those with the most NMC credit hours (mean 3.81) performed better than those with the fewest NMC credit hours (mean 3.29). The difference between the means of these two groups is statistically significant (one-way ANOVA; $q \leq 0.007$) meaning we can feel confident that our student's writing skills are improving as a result of their NMC instruction. To test this result, I compared the actual credit level (from the DataMart) of each test taker to his or her score. Indeed, those with 60 NMC credit hours and higher scored slightly better (mean 3.6) than those with less (mean 3.3).

Regression Analysis.

To ascertain the variables that could contribute to how a student scored on the essay test, I ran several regression models. The independent variables are as follows. First, the number of NMC credit hours earned by the student was used to control for age and gauge the actual amount of schooling a person had experienced. Second, a student's score on the ACT writing test, the COMPASS writing test, or the ASSET writing test was used to control for the level of skill a student had coming into NMC. Third, the students overall grade point average was used to represent the student's academic ability. In separate models, a student's grade in English 111 or whether he or she had even taken English 111 was used to represent a student's exposure to the Communications outcome. About 81% (171) of the students had completed English 111. (Only 16 students in

the sample had taken English 99, so any statistical analyses using ENG99 as a variable would be unreliable). The final models showing any statistical significance among the variables are:

Model (1): Composite Score = Credits + GPA + COMPASS + GRADEENG111

Model (2): Composite Score = Credits + GPA + COMPASS + ENG111(Y/N)

In Model (1), NMC earned credits, gpa, and COMPASS score were statistically significant, with the COMPASS score having the greatest impact (see regression coefficient table below). This means that in this model the best predictor of a student's score on the written essay test was his or her level of ability upon entering NMC. Educational experience and overall academic ability are also good predictors. In Model (2), the same three variables show statistical significance. Therefore, having taken English 111 regardless of grade is not a predictor of a student's score on the written essay test. Because a large majority of the sample had taken English 111, the variable lacks variance, which might explain why it was not significant. Therefore, three other writing intensive courses were identified as variables in the model (HST112, PLS101, and ENG112). Having taken a writing intensive course represents student exposure to the Communications Outcome. The only models that showed some level of significance are as follows:

Model (3): Composite Score = Credits + GPA + COMPASS + GradeENG112

Model (4): Composite Score = Credits + GPA + COMPASS + ENG112 OR
PLS101 OR
HST112 (Y/N)

In Model (3), NMC earned credits was statistically significant and grade in ENG112 is close, but not close enough about which to make any generalizations. In Model (4), NMC earned credits, GPA, and COMPASS score were statistically significant, but whether a student had taken one of the three courses was not. (See supporting coefficient tables in the Appendix).

What we have learned.

With regard to curriculum improvement, the vast majority of our students scored in the mid-range of the ACT rubric (between "adequate" and "competent"). Therefore, according to the rubric, to build our student's writing skills, our instruction ought to focus more pointedly on language mechanics (usage and sentence structure) and on helping students learn how to support claims with evidence. It seems their ability to organize their thoughts on paper is well proven. In fact, anecdotally, as a test administrator, I noticed students writing a rough draft of their essays or constructing an outline of their essay before putting it into the final test booklet version.

Unfortunately, it was not possible to link the CAAP writing essay scores to our own Communications Outcome rubric. The ACT rubric, comprised of four general abilities that include both writing skills and critical thinking skills, does not provide enough specific detail to be able to make the connection.

Analyzing the differences in scores among students at different educational levels is the most valuable comparison garnered from the descriptive statistics. Those students with more educational experience at NMC performed better than those with less. This fact is proven again in the regression analysis. However, the regression analysis also shows that having taken a basic-level communications course, at least those that support two levels of the Communications Outcome (some of which are writing intensive) has no bearing on the CAAP writing essay score. This is not a promising finding. It is possible that multicollinearity among the variables is a problem. However, even when GPA is removed from the models (the multicollinear culprit) the results are the same. Our students' exposure to the Communications Outcome was not a significant predictor of their CAAP composite score.

The College-Wide Assessment Team discussed and recommended that we do not use the CAAP Essay test again as a direct measure of communications ability, primarily because the ability to compare our students to a national norm is not effective in informing curriculum changes. Even so, the regression analyses show that the result of instruction in several of our writing and reading intensive courses is not having the intended impact or meeting our expectations for our general education outcomes.

If we eliminate the CAAP writing essay as an instrument of assessment, I suggest that we collect individual information on the artifacts submitted for scoring for the purpose of a more detailed analysis using inferential statistics. The student artifacts would still be chosen at random and kept anonymous but individual information would be collected. This could be another way of testing our curriculum's effectiveness.

Appendix: Supporting Data

Level	CAAP Technical Manual	
Upper-range	6 - Exceptional	These papers take a position on the issue defined in the prompt and support that position with extensive elaboration . Organization is unified and coherent. While there may be a few errors in mechanics, usage, or sentence structure, outstanding command of the language is apparent.
	5 - Superior	Same as above but with moderate elaboration
Mid-range	4 – Competent	These papers take a position on the issue defined in the prompt and support that position with some elaboration or explanation. Organization is generally clear. A competency with language is apparent, even though there may be some errors in mechanics usage, or sentence structure.
	3 – Adequate	These papers take a position on the issue defined in the prompt and support that position, but with only a little elaboration or explanation. Organization is clear enough to follow without difficulty. A control of the language is apparent, even though there maybe numerous errors in mechanics, usage, or sentence structure.
Lower-range	2 - Weak	While these papers take a position on the issue defined in the prompt, they may show significant problems in one or more of several areas, making the writer’s ideas often difficult to follow: support may be extremely minimal; organization may lack clear movement or connectedness; or there maybe a pattern of errors in mechanics, usage, or sentence structure that significantly interferes with understanding the writer’s ideas.
	1 - Inadequate	These papers show a failed attempt to engage the issue defined in the prompt, lack support, or have problems with organization or language so severe as to make the writer’s ideas very difficult to follow.

Ethnicity	Frequency	Percentage
African American/Black	2	0.9%
American Indian	2	0.9%
White	178	83.2%
Mexican American	3	1.4%
Asian	2	0.9%
Prefer not to respond	22	10.3%
No response	5	2.3%
Total	214	100.0%

Gender	Frequency	Percentage
Male	86	40.4%
Female	127	59.6%
Total	213	100.0%

Age	Frequency	Percentage
18 and under	32	15.0%
19-20	77	36.0%
21-25	59	27.6%
26-30	14	6.5%
31-39	15	7.0%
40 and older	17	7.9%
Total	214	100.0%

Credit Level	Frequency	Percentage
Freshman	84	39.3%
Sophomore	74	34.6%
Junior	35	16.4%
Senior	12	5.6%
Other	8	3.7%
No Response	1	0.5%
Total	214	100.0%

Cumulative GPA	Frequency	Percentage
Below 2.0	7	3.3%
2.01-2.50	34	15.9%
2.51-3.0	52	24.3%
3.01-3.5	76	35.5%
3.51 and above	42	19.6%
No response	3	1.4%
Total	214	100.0%

Educational level	Mean	N	Std. Deviation
Freshman	3.2946	84	.45615
Sophomore	3.2804	74	.68491
Junior	3.4286	35	.58965
Senior	3.8125	12	.48996
Other	3.7500	8	.29881
Total	3.3580	213	.57980

DATAMART EDU LEVEL	Mean	N	Std. Deviation
0-29 credits	3.2677	99	.60319
30-59 credits	3.3320	64	.54382
60-89 credits	3.6216	37	.57922
90 credits and above	3.5000	10	.33333
Total	3.3607	210	.58280

Educational Major	Mean	N	Std. Deviation
Undecided	3.3421	38	.49121
Architecture	3.4167	3	.62915
Biological Sciences	3.8750	2	.17678
Business	3.1250	24	.80081
Office Management	3.0000	1	.
Communications	4.2500	2	1.06066
Community Services	3.2500	11	.60208
Computer and Info Sciences	3.1563	8	.44194
General Studies	3.0833	3	.14434
Education	3.5366	41	.54360
Engineering	3.1250	2	.17678
Fine Arts	3.2500	4	.28868
Health Professions	3.1786	14	.43222
Home Economics	3.4792	24	.48295
Letters	3.1667	3	1.04083
Physical Sciences	3.4375	4	.51539
Social Sciences	3.5455	22	.57028
Trade and Industrial	3.1250	2	.17678
Total	3.3726	208	.57262

MODEL 1	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
REGRESSION DV= COMPOSITE SCORE (Constant)	2.068	.223		9.279	.000
NMC_EARN_HRS	.005	.002	.209	2.203	.030
OVERALL_GPA	.227	.100	.267	2.270	.025
ENG 111 GRADE	.011	.080	.016	.135	.893
COMPASS_WRITING	.005	.002	.237	2.608	.011

MODEL 2	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
REGRESSION DV= COMPOSITE SCORE	B	Std. Error	Beta		
(Constant)	2.045	.220		9.298	.000
NMC_EARN_HRS	.005	.002	.194	2.203	.029
OVERALL_GPA	.152	.073	.180	2.072	.040
REGRAD11	-.045	.128	-.030	-.349	.728
COMPASS_WRITING	.010	.002	.381	4.624	.000

MODEL 3	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
REGRESSION DV= COMPOSITE SCORE	B	Std. Error	Beta		
(Constant)	1.747	.311		5.612	.000
NMC_EARN_HRS	.007	.003	.247	2.379	.020
OVERALL_GPA	.197	.130	.210	1.511	.135
COMPASS_WRITING	.002	.003	.097	.952	.344
ENG112	.183	.098	.253	1.871	.065

MODEL 4	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
REGRESSION DV= COMPOSITE SCORE	B	Std. Error	Beta		
(Constant)	2.004	.204		9.807	.000
NMC_EARN_HRS	.006	.003	.204	2.165	.032
OVERALL_GPA	.155	.074	.184	2.092	.038
COMPASS_WRITING	.010	.002	.395	4.906	.000
RECODCOM	-.055	.122	-.042	-.446	.656