

Cluster Three: The Natural World  
Cluster Assessment Results and Interpretation

*Spring 2006 Assessment Results*

Center for Assessment and Research Studies  
James Madison University

## Spring 2006, Cluster 3 Assessment Report

### **Executive Summary**

Nw-8 test results from 501 students with between 45-70 credit hours were analyzed using three assessment strategies. Scores on the NW-8 test were higher for students who had taken more JMU Cluster 3 courses. This occurred regardless of whether the students had AP/Transfer credits for Cluster 3 courses. Scores on the NW-8 were positively correlated with Cluster 3 course grades for the majority of Cluster 3 courses. A few other courses had near zero correlations or negative correlations. Students who completed their Cluster 3 requirement showed higher scores than first-year students and students who had not completed cluster 3 requirements.

## Spring 2006, Cluster 3 Assessment Report

### Introduction

This report describes the results of the Spring 2006 Assessment Day administration of the Natural World test version 8 (NW-8). The NW-8 assessment was designed by faculty and assessment specialists at JMU to measure the objectives of Cluster 3 (CL3), the Natural World segment of JMU's general education program. Spring 2006 was the first time the NW-8 was administered to second-year JMU students. This version reflects revisions resulting from extensive validity studies on previous versions of the Natural World instrument. This report contains the results of statistical analyses that describe the relationship between the CL3 program and NW-8 scores. These analyses lend support to the perception that students are learning as a result of participation in CL3 program. In addition, the report will present newly available validity evidence for this version of the NW test.

### Sample and Data collection procedures

Five-hundred twenty-six randomly selected students with 45-70 credits completed the pencil and paper version of the NW-8 assessment on Assessment Day (February 14, 2006). Since typical JMU assessment day procedures were followed, the quality of the data collected can be regarded as comparable to that of previous administrations. If students reported an invalid student ID or failed to provide a response to more than half of the test items, their test scores were excluded from the analyses. Thus, the final sample size for the analysis was 510 students, which was comprised of 330 females (65%) and 180 males. The sample is smaller than previous year's assessment samples due to logistical constraints of test administration across the university and other unforeseen factors. In future years, sample sizes for CL3 assessment should return to normal (closer to 1,000 students). While the majority of the sample was sophomore-level, juniors were represented as well (36%). The percentages of ethnic backgrounds of the students in the sample (83% White, 5% Asian, 2.5% Black, 2.5% Hispanic, and 6% other) roughly approximated those of the overall JMU population.

### Reliability and Scores

Test scores were computed by assigning 1 point for each item answered correctly and 0 points for incorrect or unanswered items then summing across items included in the subtest. Scores are reported as both raw number of points and percent correct scores. Table 1 shows the internal-consistency reliability estimate (Cronbach's coefficient alpha, represented as  $\alpha$  in the table) for the total test and each subtest. Cronbach's  $\alpha$  refers to the degree that scores are free from measurement random error; in other words the extent to which item responses are systematic. Notice that the table presents the scores and reliabilities for two separate NW-8 test blueprints, the original one and the revised one. In 2006, CL3 faculty and assessment practitioners conducted extensive content validity studies on the NW-8 test, resulting in a new test blueprint which differs from the original one. The revised blueprint has more items assigned to more than one objective (because these items were recognized to measure multiple objectives). Furthermore, no items were mapped to Objective 2 because it was found to be too global to which to assign specific items. For all objectives except objectives 1, subscale reliability improved when implementing the revised test blueprint. Conventionally,  $\alpha$  values above .70 are considered adequate for program evaluation or research. Thus, total test score, Quantitative Reasoning score, and Objective 7 subscore have adequate reliability to interpret means and scores, where as the other objective scales do not.

**Table 1***Means, standard deviations and reliability estimates by CL3 objective*

<b>Cluster 3 Objective</b>	<b>Blueprint</b>	<b>Included items</b>	<b>Results</b>
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	Original	9 items: 9, 15, 20, 34-38, 41 (14% of total test)	M = 7.38 (82% correct); SD = 1.59; $\alpha = .55$
	Revised	10 items: 2, 5, 9, 15, 20, 35-38, 41 (15% of total test)	M = 8.05 (80% correct); SD = 1.61; $\alpha = .48$
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	Original	9 items: 4-7, 19, 39, 40, 43, 56 (14% of total test)	M = 6.64 (74% correct); SD = 1.68; $\alpha = .49$
	Revised	0 items	
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	Original	9 items: 1, 16-18, 44-48 (14% of total test)	M = 5.68 (63% correct); SD = 1.90; $\alpha = .50$
	Revised	11 items: 1, 16-18, 24, 42, 44-48 (17% of test)	M = 7.15 (65% correct); SD = 2.18; $\alpha = .55$
4. Illustrate the interdependence between developments in science and social and ethical issues.	Original	10 items: 2, 21-24, 42, 58-61 (15% of total test)	M = 7.95 (80% correct); SD = 1.55; $\alpha = .43$
	Revised	13 items: 2, 16-18, 21-24, 42, 58-61 (20% of test)	M = 9.96 (77% correct); SD = 1.94; $\alpha = .49$
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	Original	13 items: 8, 10-14, 27-29, 53-55, 62 (20% of total test)	M = 7.54 (58% correct); SD = 2.13; $\alpha = .53$
	Revised	13 items: 4, 7, 8, 19, 27-29, 43, 53-55, 57, 62 (20% of total test)	M = 7.57 (58% correct); SD = 2.14; $\alpha = .56$
6. Discriminate between association and causation, and identify the types of evidence used to establish causation.	Original	7 items: 3, 25, 30-33, 57 (11% of total test)	M = 5.38 (77% correct); SD = 1.54; $\alpha = .57$
	Revised	18 items: 3, 6, 10-14, 25, 30-33, 40, 53-57 (28% of total test)	M = 12.69 (71% correct); SD = 2.90; $\alpha = .67$
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	Original	10 items: 10-14, 26, 49-52 (15% of total test)	M = 8.10 (81% correct); SD = 1.78; $\alpha = .61$
	Revised	19 items: 5, 6, 9-14, 20, 25, 26, 39-41, 49-52, 56 (29% of test)	M = 15.10 (79% correct); SD = 2.92; $\alpha = .70$
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	Original	7 items: 35-38, 63-65 (11% of total test)	M = 5.85 (84% correct); SD = 1.34; $\alpha = .55$
	Revised	18 items: 2, 15, 21-23, 34-38, 42, 57, 59-61, 63-65 (28% of test)	M = 14.86 (83% correct); SD = 2.51; $\alpha = .64$
Quantitative Reasoning (Objectives 5 and 6)	Original	20 items: 3, 8, 10-14, 25, 27-33, 53-55, 57, 62 (31% of test)	M = 12.92 (65% correct); SD = 3.09; $\alpha = .67$
	Revised	27 items: 3, 4, 6-8, 10-14, 19, 25, 27-33, 40, 43, 53-57, 62 (42% of test)	M = 18.15 (67% correct); SD = 3.99; $\alpha = .73$
<b>Total Test</b>		65 items M = 47.56 (73% correct) SD = 8.10 $\alpha = .85$	

## Assessment Strategy I: Differences

*Hypothesis:* Do students who have completed more CL3 courses perform better on the NW-8 assessment?

*Prediction:* Students who have completed a greater number of CL3 courses should score higher on the NW-8 test than students who have taken fewer courses in the cluster.

To investigate this question, we grouped students according to how many CL3 courses they have completed and compared the mean total NW-8 score across groups. We looked at this pattern in two samples: all students (N=510), and excluding students who have Advanced Placement (AP) exam or transfer (TR) credits related to CL3 material (n=335). The second sample was considered because it is plausible that AP and TR experiences outside of the CL3 program may impact students' scores on the NW-8 test. Excluding students who have this alternative influence clarifies the effect of the CL3 program.

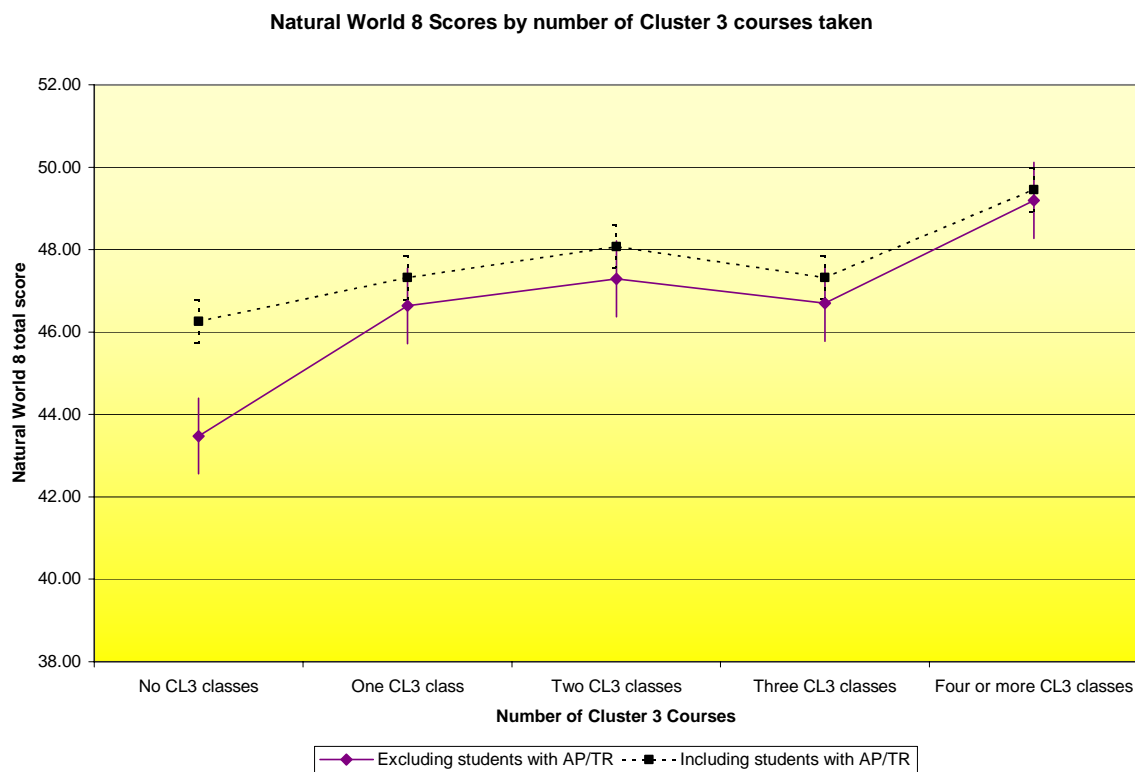
Table 2 displays the means for each of the two samples. Figure 1 displays the pattern of means graphically. Clearly, there is an upward trend in NW-8 scores as students take more CL3 courses. This trend is statistically significant when AP and TR credits are controlled for ( $r = .112$ ,  $p = .04$ ). Also, as expected, students who have completed AP or TR credits related to CL3 score higher on the NW-8 test than students who have not.

Table 2

*Means for students without AP or TR credits, and for all students*

Number of JMU CL3 Courses	n	Raw Score Mean	Raw Score SD	Percent Correct Mean	Percent Correct SD
<i>Excluding Students who have AP or TR Credits</i>					
0	21	43.48	6.86	66.86	10.61
1	131	46.63	7.64	71.74	11.78
2	103	47.29	8.27	72.80	12.84
3	49	46.69	8.11	71.80	12.40
4 or more	31	49.19	6.71	75.87	10.33
Total	335	46.88	7.83	72.15	12.09
<i>Including Students who have AP or TR Credits</i>					
0	78	46.26	9.29	71.18	14.36
1	176	47.32	7.64	72.80	11.76
2	147	48.07	8.21	73.99	12.69
3	65	47.32	8.02	72.83	12.27
4 or more	44	49.45	7.21	76.23	11.02
Total	510	47.56	8.10	73.19	12.48

Figure 1.



*Note.* Error bars in Figure 1 represent the expected fluctuation in means due to sampling variability (i.e. the standard error).

## Assessment Strategy II: Relationships

*Hypothesis:* What is the relationship between course grades and scores on the NW-8 assessment?

*Prediction:* A positive relationship should exist between course grades in CL3 courses and scores on the NW-8 test (Total scores and Quantitative Reasoning scores).

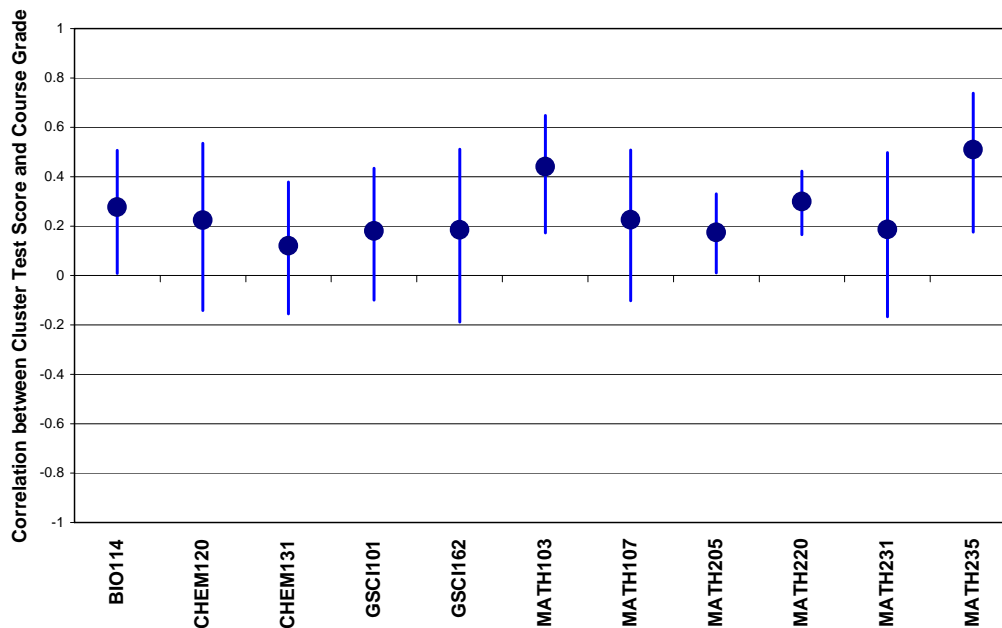
To investigate whether high grades in CL3 courses corresponds with higher scores on the CL3 assessment test, we correlated course grades with NW-8 grades for each course for which at least 30 students in the sample had received a grade. Course grades were correlated with overall NW-8 scores and with QR subtest scores; these correlations are reported in Table 3. Correlations can range from -1.0 to 1.0. Correlations close to 0 indicate a weak relationship, while correlations closer to 1 indicate a strong relationship between test scores and course grades. Negative numbers indicate an inverse relationship: those who scored high on the test had lower course grades, and those with low test scores had higher CL3 course grades. With the exception of GSCI104, all correlations were positive, as expected. In this context, correlations of about .2 seem good and correlations of .4 seem fairly high.

Table 3.  
*Correlations between Grades and NW-8 scores*

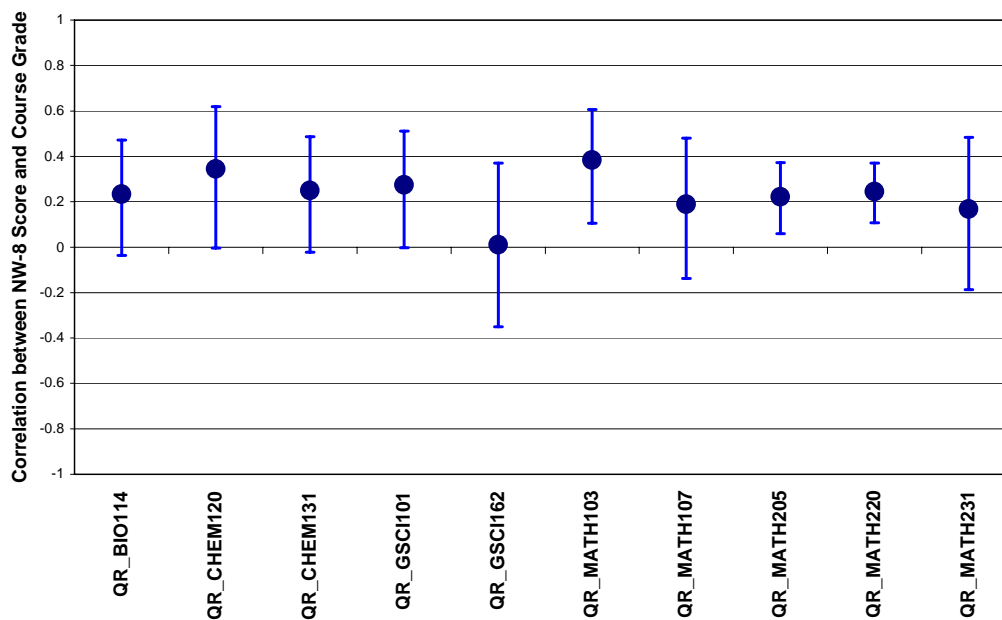
COURSE	N	NW	NW R <sup>2</sup>	QR	QR R <sup>2</sup>
BIO114	54	0.276	0.08	0.233	0.05
BIO270	16				
CHEM120	32	0.412	0.17	0.344	0.12
CHEM131	53	0.189	0.04	0.249	0.06
ANTH196	7				
GEOL110	16				
GEOL211	0				
GISAT112	15				
GISAT113	11				
GSCI101	51	0.181	0.03	0.274	0.07
GSCI102	14				
GSCI103	7				
GSCI104	24	-0.035	0.00	-0.044	0.00
GSCI116	4				
GSCI115	3				
GSCI121	7				
GSCI122	1				
GSCI162	30	0.185	0.03	0.011	0.00
GSCI163	9				
GSCI164	8				
GSCI165	11				
MATH103	46	0.441	0.19	0.384	0.15
MATH107	38	0.230	0.05	0.189	0.04
MATH205	143	0.205	0.04	0.222	0.05
MATH220	196	0.231	0.05	0.244	0.06
MATH231	33	0.187	0.04	0.167	0.03
MATH235	29				
PHYS120	0				
PHYS140	0				
PHYS240	4				

The graphs below plot the  $r$  values and illustrates the instability in the correlations due to sampling error. The error bands show the 95% confidence band around the correlation estimates.

### Natural World -8 scores with course grades



### Quantitative Reasoning-8 scores with course grades



### Assessment Strategy III: Change over time

*Hypothesis:* Do students who participate in the CL3 program improve their scores more over time than students who do not participate in the CL3 program?

*Prediction:* Scores on the NW-8 test should increase more for students who have had exposure to CL3 courses than for students who have not had exposure to CL3 courses.

This question could not be addressed in this report because longitudinal, value-added data was not available for this cohort using version 8 of the NW test. These students had completed version 7 of the NW test as first-year students. The two versions are too dissimilar to compare scores. However, we can show how first-year JMU students from other cohorts have performed on the NW-8 exam which may suggest how much change has occurred for students overall. Table 4 shows that JMU sophomores perform about 12 percentage points higher on the NW-8 than JMU first-year students in Fall 2006. This difference is statistically significant  $t= 13.177$ ,  $df= 1961$ ,  $p < .001$ ,  $d= .68$ . Furthermore, the difference is considered medium-large according to Cohen's  $d$  effect size benchmarks. When first-year JMU students are compared to JMU students who have completed the cluster 3 requirement, then the mean difference is still significant and is considered very large,  $t= 6.389$ ,  $df= 1503$ ,  $p < .001$ ,  $d= .90$ .

Table 4  
*Total scores on the NW-8 assessment*

Group	n	Raw Score Mean	Raw Score SD	Percent Correct Mean	Percent Correct SD
JMU first-year student	1,453	43.47	7.31	67	10.61
All JMU sophomores	501	46.63	7.64	71.74	11.78
Cluster3 completers	52	50.06	7.27	77.01	11.19
Only Non-completers	103	47.30	8.15	72.73	12.53

**Assessment Strategy IV. Meeting a standard (expectation)**

*Do students meet faculty expectations?* A substantial number of students who have completed the Cluster should meet the competency or academic standard.

This assessment strategy does not apply to Cluster 3 because a standard for the NW-8 is currently slated to be established in 2007.

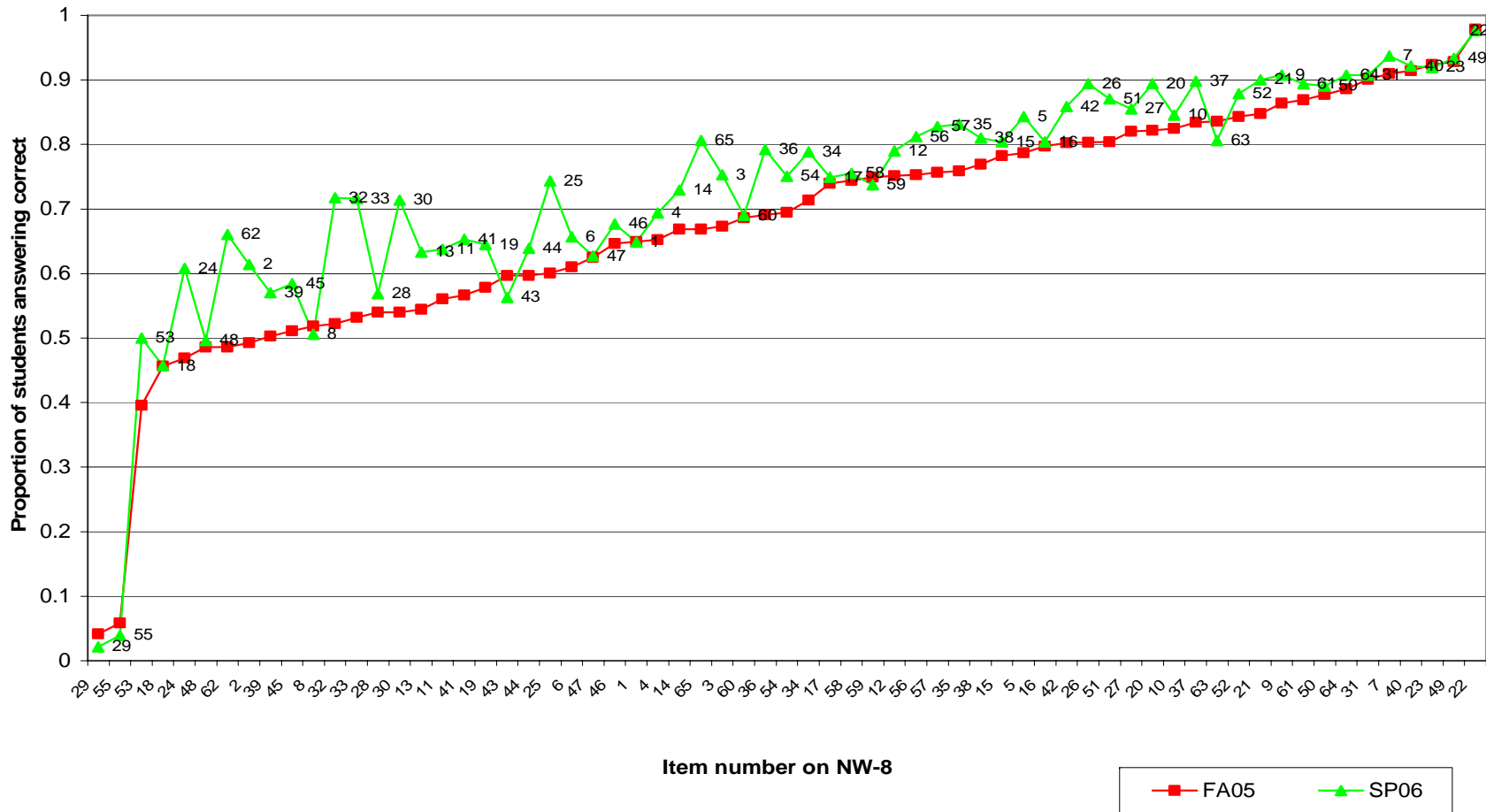
**Conclusions/Recommendations**

Overall, results from the current year's Cluster 3 assessment were as expected. All predictions of student performance were supported. A fourth type of assessment strategy (Meeting a standard) will be possible when a standard has been set for the NW-8 test. Item analysis results (Shown in Appendix II) suggest that items 29 and 55 may need to be reviewed because they appear to be much more difficult for students than the other items as a whole. The results of this semester's assessment were limited because the sample size was smaller than usual. Furthermore, pre-post comparison of scores on the NW-8 could not be conducted because the NW version 8 had not been created when these students were entering first-years.

**Appendix I**

The plot below shows the proportion of students answering each NW-8 item correctly at pre-test and post-test. More students answer the items correctly in Spring 2007 than did in Fall 2005. However, some items do not show much change, while other items show more change from Fall to Spring.

**Proportion of Students answering each item correctly as first-year students and as students with 45-70 credits**



## Appendix II

### NW-8 Validity Studies

One form of validity evidence that can be collected is the correlation of NW-8 scores with another external variable. If the two variables relate in expected ways, support is lent for the notion that the test is measuring what it is supposed to measure. Two external variables that we expect to be positively correlated with NW-8 scores are the Quantitative component of the Scholastic Aptitude Tests (SAT) and the number of Advanced Placement (AP) exams passed in the areas of math and science. Specifically, scores on the NW-8 test should be moderately positively correlated with scores on the Quantitative component of the Scholastic Aptitude Tests (SAT) because both tests purport to measure quantitative reasoning to some degree. In addition, scores on the NW-8 are expected to be related to whether students have passed AP exams in math and science related subjects. Students who have passed more AP exams in math and science should score higher on the NW-8 test.

Table 5 presents the correlation between SAT and NW-8 scores, as well as the correlation between number of AP exams in math and science passed and NW-8 scores. Correlations are provided for both the total NW-8 score and the QR reasoning subtest score. All correlations were significant and can be considered moderately positively correlated.

Table 5  
*Correlations with NW-8*

	<b>Total score NW test</b>	<b>Quantitative Reasoning subscore</b>
<b>SAT Quantitative score</b>	.423*	.401*
<b>No. of AP tests passed</b>	.269*	.229*

\*significant at the  $p < .05$  level