Cluster Three Annual Report
2007-2008

A. Introduction: Overall (*Summary of the Cluster…)

1. Year-End Executive Summary

There were two areas of significant accomplishment in Cluster Three this year. First, the Cluster Coordinator developed a model to predict the number of current and future spaces needed in the cluster to ensure students complete the cluster in a timely fashion. This model was reviewed by the Cluster Committee and department heads and deans in the cluster, and the queries needed to obtain data to monitor enrollments have been written and are ready for use.

Second, the cluster made substantial progress related to assessment. Significant improvements were made to the assessment instrument, resulting in a new, ninth version. The Cluster Committee developed a new set of assessment goals aimed at making assessment more useful to faculty teaching courses. The committee and CARS develop a method for comparing faculty expectations and student performance on individual items, items representing individual objectives, items representing quantitative reasoning, and the entire instrument. Results from administration of the new instrument to entering students and sophomores/juniors showed that student performance increased with increasing number of Cluster Three courses, and with the amount of AP credit in Cluster Three brought to JMU. The percentage of students exceeding faculty expectations was greater for sophomores/juniors than for freshman for all objectives, as well as for the total score.

B. Cluster Objectives

1. List cluster learning objectives

- Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudoscience.
- Use theories and models as unifying principles that help us understand natural phenomena and make predictions.
- Recognize the interdependence of applied research, basic research, and technology, and how they affect society.
- Illustrate the interdependence between developments in science and social and ethical issues.
- Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.
- Discriminate between association and causation, and identify the types of evidence used to establish causation.
- Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.
- Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.

2. Has any change been made to the cluster objectives since the last cluster review? If so, provide information on the changes.

   No changes were made this year.

3. Are there plans to add, delete, and/or modify the cluster objectives?

   There are no plans to add, delete, and/or modify the cluster objectives.

C. Unit Accomplishments – Planned Cluster Activities

1. List and state purpose of cluster activities
   There were three planned activities of the cluster this year:
   a. Implement efforts to reduce the level of students taking Cluster Three courses.
      Providing sufficient spaces in Cluster Three continues to be a challenge, and many students delay their completion of the cluster because they are unable to register for courses until they obtain sufficient priority based on accumulated credit hours. Reducing the level of students taking Cluster Three courses requires determining which courses need additional spaces and whether students are completing the cluster in a timely fashion. The following activities would provide the ability to predict needs and allocate teaching FTEs to provide sufficient spaces in the cluster in an efficient manner:
      - Develop a model for predicting the number of spaces required in Cluster Three over 1-, 2-, 3- and 4-year planning periods.
      - Determine the appropriate contribution of each department based on numbers of instructors, majors, and courses serving students other than majors or GenEd
      - Develop a procedure for communicating to departments expected number of spaces for an academic year, given current enrollment at JMU.
      - Gather and analyze data to determine whether adequate spaces are being offered:
        - the average level of students (i.e., freshman, sophomore, etc.) in current courses.
        - the proportion of students completing their Cluster Three requirements by various times (e.g., second semester junior year).
        - the average number of Cluster Three courses taken at JMU by students during their time here
      - Share strategies to reduce the level of students taking Cluster Three courses with cluster department heads and deans during annual meetings with these individuals.

   b. Work with departments to standardize and update how AP credit is counted toward Cluster Three
      There are a number of inconsistencies in the way AP credit is awarded across the cluster due to the change in the structure of the cluster two years ago. These

inconsistencies need to be resolved to facilitate the correct awarding of AP credit in the cluster.

c. Cluster Three Assessment
There were two goals for work on assessment in the cluster:
- Conduct a standard setting session to establish a benchmark for performance on the Cluster Three Assessment instrument.
- Review and revise assessment goals for the cluster.
The assessment goals of the cluster had not been reviewed for a number of years. A coherent set of goals generated by the cluster faculty is essential if assessment is to provide faculty with assessment data they can use to improve cluster courses and the cluster's curriculum.

2. How were these cluster activities implemented
a. Implement efforts to reduce the level of students taking Cluster Three courses.

The Cluster Three Coordinator completed the bulk of the work related to this activity, receiving feedback and input from the Cluster Three Committee, Cluster Three department heads and deans, and the IDLS science concentration coordinator. The Cluster Coordinator also enlisted the help of Peter DeSmit, of JMU's Information Systems, to develop the queries that would provide data on when students complete Cluster Three courses.

b. Work with departments to standardize and update how AP credit is counted toward Cluster Three.

In annual meetings with department heads, the Cluster Three Coordinator requested that departments review AP credit.

c. Cluster Three Assessment

<u>Standard Setting</u>
On June 22, 2007 CARS hosted a standard setting session with Cluster Three faculty to establish a benchmark for performance on the Cluster Three Assessment instrument.

<u>Assessment Goals</u>
The Cluster Three Committee representatives solicited input from their departments about what research questions they would like to address as part of Cluster Three Assessment. The committee compiled those suggestions and adopted the three most commonly mentioned goals as the new Cluster Three assessment goals.

3. State the outcome of each cluster activity
   a. Implement efforts to reduce the level of students taking Cluster Three courses.

      The Cluster Three Coordinator developed a proposal that detailed a predictive model of the number of spaces needed in the cluster, a method for determining how those spaces should be allocated across departments, and a data collection system for monitoring enrollments and backlogs. The Coordinator shared the proposal with the Cluster Three Committee, and Cluster Three deans and department heads. He incorporated feedback from these constituents and finalized the proposal for implementation. Peter DeSmit developed the queries necessary to gather the essential data to monitor the level of students in courses.

   b. Work with departments to standardize and update how AP credit is counted toward Cluster Three

      The Cluster Coordinator has not received any response from departments on standardizing and updating AP credit.

   c. Cluster Three Assessment
      The standards setting session was completed but no standard was set because the faculty participating in the session felt that a number of items on the instrument were problematic, thus making it very difficult to produce a valid cutoff score for "meets expectations." The assessment was significantly revised as a result of this session.

      The Cluster Three Committee decided on the following three assessment goals:
      1. What objectives are being mastered and which are not?
      2. What common misconceptions do students have about the concepts embodied by the objectives?
      3. Do Cluster Three courses change students attitudes toward math and science and student perception about the relevance of math and science to their lives and students' confidence in using math and science on an everyday basis?

      The Cluster Three Committee made significant progress addressing Item 1 (which objectives are being mastered). Thirty-four faculty from across the cluster rated what percentage of students completing the cluster would answer correctly each item on the assessment instrument. This process generates a faculty expectation that is then compared to student performance for both incoming freshman and sophomore/juniors assessed on Assessment Day. Comparisons can be made for individual items, objectives, or the entire instrument.

iv. What obstacles prevented completion of the cluster activities

The Cluster Coordinator was away from campus for the second half of Spring 2008, which prevented completion of data gathering and analysis for the model predicting needs in the cluster.

v. *Summary of Cluster Activity/Highlight significant accomplishments

During the past academic year, the Cluster Coordinator developed a model to predict the number of current and future spaces needed in the cluster to ensure students complete the cluster in a timely fashion. This model was reviewed by the Cluster Committee and department heads and deans in the cluster. The queries needed to obtain data to monitor enrollments have been written and are ready for use.

The Cluster Three Committee and CARS have made significant progress with assessment this year, having developed three new goals for cluster-wide assessment. Thirty-four faculty participated in quantifying faculty expectations for student performance on the assessment instrument, and the committee and CARS develop a method for comparing faculty expectations and student performance on individual items, items representing individual objectives, items representing quantitative reasoning, and the entire instrument. The instrument was significantly revised, resulting in the ninth version of the instrument.

D. Assessment Report

1. Assessment Progress Template (see Appendix A)
   1. Mapping cluster objectives to curriculum
   2. Mapping cluster objectives to assessment methods
   3. Matrix of six key elements in cluster assessment template
2. Include full report as an attachment
   a. Fall 2007 Report (Appendix B)
   b. Spring 2008 Report (Appendix C)

3. *Summary Page

From the Fall 2007 report:

A total of 1,408 entering freshman produced useable tests, with 71% administered in paper-and-pencil format, and 29% via computer. The test exhibited high reliability ($\alpha = .64$) for Quantitative Reasoning (QR) and the total test ($\alpha = .78$), which is how Scientific Reasoning (SR) is conceptualized; these reliabilities are adequate for program evaluation purposes and comparable to those observed for earlier entering first-year cohorts at JMU. Reliability estimates obtained for computer-based administration were slightly lower than those for the paper and pencil format. Entering student performances for QR were at about 62% correct and 64% correct for the total SR score. Students entering the university with Advanced Placement (AP) or Dual Enrollment college credits tended to score 4-10% higher than students entering without relevant college-level credits. Additional analyses produced provide evidence that both AP and Dual Enrollment credits contribute to higher scores on both QR and SR.

From the Spring 2008 report:

A total of 1,020 students with 45-70 credits produced useable tests on Assessment Day testing, with 90% administered in paper-and-pencil format, and 10% via computer. The test exhibited high reliability for QR (α = .66) and SR (α = .82); these reliabilities are adequate for program evaluation purposes. Because this is the first time this version of the instrument has been administered to sophomores/juniors, no pre- and post-test comparisons can be made. Entering student performances for QR were at about 68% correct and 71% correct for the total SR score. Student performance increased with increasing number of Cluster Three courses, and with the amount of AP credit in Cluster Three brought to JMU. The percentage of students exceeding faculty expectations was greater for sophomores/juniors than for freshman for all objectives, as well as QR and SR (entire test).

E. SCHEV Report (if applicable)

Quantitative Reasoning

1. Definition of terms

1.1: James Madison University has established five General Education Clusters of learning objectives required for all students. Cluster Three: The Natural World is comprised of learning objectives most closely associated with Quantitative Reasoning. Two of the goals and objectives within Cluster Three define quantitative reasoning competency for JMU students and faculty. Students competent in quantitative reasoning will be able to: 1) Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena and 2) Discriminate between association and causation, and identify the types of evidence used to establish causation.

1.2: JMU will report the percentage of students meeting faculty expectations for acceptable student performance. The standard setting session for the 8th version of the Quantitative Reasoning assessment has been scheduled for July, 2008. As part of the Bookmark standard setting procedure we employ, empirical student performance [impact] data from spring 2007 QR-8 administrations will be provided to Cluster 3 faculty judges.

1.3: Students competent in quantitative reasoning will be able to:

- Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.
- Discriminate between association and causation, and identify the types of evidence used to establish causation.

These learning objectives provide the framework upon which courses are reviewed, approved, and assessed for the quantitative reasoning component of the Natural World.

2. Methodology

2.1 To demonstrate competency James Madison University will use a pre-post value added analytic approach. JMU’s Assessment Day data collection strategy has been in place for 21 years and provides data on entering students (fall) and those that have completed 45 to 70 credit hours (spring).

2.2 By assessing students as they enter the University and again after they have earned 45 to 70 credits, JMU is in the position to answer questions regarding differences in students as they enter the University (including those with transfer, Advance Placement, dual-enrollment credits). Also, given that some students in the 45 to 70 credit range have completed their quantitative reasoning course requirements and others have not, we can compare the performance of students at various stages of completion in an area and examine differences. This process also allows JMU to examine pre-post differences.

2.3 The 27-item instrument employed is our locally developed Quantitative Reasoning instrument (QR-8), which is now in its 8th version. This instrument, in combination with our Scientific Reasoning (SR-8) test, is the focus of a newly funded NSF project, and has shown appropriate psychometric characteristics at JMU and beyond. Cronbach alpha reliability estimates are typically .59 for first-year students and .75 for sophomore JMU samples.

2.4 Two large, representative, and random samples of students are selected for assessment in the Natural World during two Assessment Days: 1) Entering-first-year students during August orientation (approximately 1,000 students); and 2) a sample of students (approximately 900 students) at the midpoint of their undergraduate academic careers [45-70 credit hours] during the annual February Assessment Day.

2.5 During JMU Assessment Day activities, students are randomly assigned to testing rooms based on the last two digits of the students' student identification number. Students are then assigned by room to complete cognitive (such as *quantitative reasoning*) and non-cognitive (such as *attitudes toward learning)* assessments. This random assignment allows JMU to assess students in all general education areas every year. Assessments are administered using standardized testing procedures and trained proctors.

2.6 Among the analytical methods used to examine the pre-test scores of entering students on Quantitative Reasoning, JMU employs the following:

- Differences—Students who enter the university and are transferring in no credits, those who transfer is AP credits and those transferring in dual-enrollment credits.
- Competency—The percentage of the students who achieve the standard or expectation established by JMU faculty as entering students.

2.7 Among the analytical methods used to assess the impact (pre-post) of Cluster Three courses in achieving Quantitative Reasoning, JMU employs the following:

- Differences—Students who have completed all Natural World course requirements should perform better than students who have taken only one or two courses related to quantitative reasoning.
- Change or Value-Added—Sophomore and junior-level students should perform better on the Quantitative Reasoning test than entering first-year students. Further, students who have completed their Natural World course requirements should perform better on the Quantitative Reasoning test than they did as entering first-year students [Value-Added].

- Competency—The percentage of the students who have completed their Cluster Three quantitative reasoning coursework that achieve the standard or expectation established by JMU faculty. The performance level associated with competency will be defined at a July, 2008 standard setting session involving JMU and Virginia Tech faculty.

2.8 Establishing Change or Value-Added:
Because students are assigned to testing locations on the basis of the last two digits of their JMU ID and all entering and sophomore/juniors are required to participate in their respective Assessment Day, JMU can provide both cross-sectional and true repeated-measure samples. Our first true repeated measures sophomore sample using QR-8 was collected in spring 2007.

The magnitude of proficiency change over time will be calculated by administering the QR-8 at two time points: pretest: August Assessment Day; and posttest: February Assessment Day. The average pretest and posttest scores of students who have completed the relevant general education coursework will be used. This difference score will be standardized by dividing the difference by the pretest standard deviation to enhance interpretation and reporting. Quoting from the *General Education Assessment at JMU Instrument Categories* document [which is available for download at http://www.jmu.edu/assessment/JMUAssess/Gened/GenEd%20Assessment_Instrument%20Categories.pdf ,"The specific standardized mean difference index employed is Cohen's d defined as follows:

$$d = \frac{M_{post} - M_{pre}}{SD_{pre}}$$

with Mpost representing the average posttest raw score, Mpre the average pretest raw score and SDpre the standard deviation of the pretest raw scores. The interpretation of the resulting d is best conveyed using an example: if d = 0.5, the average posttest scores are half a standard deviation unit higher than average pretest scores. In addition to Cohen's d and the descriptive statistics for the pretest and posttest scores, the correlation between pretest and posttest scores should also be reported so that meta-analytic studies that combine the standardized mean difference across cohorts and measures can be pursued." (p. 3). Examinee motivation must be a consideration for appropriate interpretation of performances emanating from testing contexts without personal consequences for individual examinees. Continued inference validation research is being conducted at JMU to address these issues; thus far, our cumulative evidence suggests student motivation is sufficiently high for most students.

3. Process evaluation

3.1 Annually, reports are generated and distributed to the Cluster Steering Committees of all five clusters, including the cluster that reviews quantitative reasoning. In addition, each JMU Cluster is reviewed on a 5-year cycle where program efficacy is carefully monitored and data use for program improvement is expected.

3.2 The assessment process at JMU is fully integrated with an institutional reporting system by which academic programs complete an annual Assessment Progress Template

(APT; see http://www.jmu.edu/assessment/JMUAssess/apt.html) as part of the departmental annual report. All programs also complete Academic Program Review (APR) on a six-year cycle. The newly revised guidelines can be reviewed at http://www.jmu.edu/acadaffairs/wm_library/Revised%20Guidelines%20of%2006-05-07.pdf). Academic programs are expected to report five years of assessment data as part of the APR process. All General Education program clusters are reviewed on a five year cycle, and the General Education program as a whole participates in the University's Academic Program Review process. JMU's 2007 Six Year Plan submitted to SCHEV includes several goals closely associated with quantitative reasoning, the continued adoption and full implementation of the APT, strengthening the relationship of the APT to the APR process and decision making (see JMU goals 4.1, 4.2, 4.3, and 4.7). In addition, Goal 4 of JMU's Defining Characteristics for 2006-2012 states the institution's commitment to "Establish and assess optimum competencies for the General Education program" of which quantitative reasoning is a member.

3.3 James Madison University's assessment expenditures include funding of the Center for Assessment and Research Studies, as well as providing support across campus for program assessment. Estimated annual cost is $565,755 for the assessment related component of services. It should be noted that this annual cost includes assessment for JMU's five clusters of general education, student affairs assessment and program assessment. Estimated cost for a single cluster, in this case quantitative reasoning, is $40,411 annually.

4. Data presentation

4.1 Several challenges were encountered and overcome while working toward delivery of data collection and analysis. The Standard Setting scheduled for July 2008 was conducted but not successful. Several faculty members were recruited from JMU, VCCS and Virginia Tech to provide judgments to develop standards for student performance. Several JMU faculty members expressed concern related to the composition and wording of items comprising the instrument. As a result, the QR-8 and SR-8 instruments were subjected to complete review. Fortunately, JMU's recent NSF grant funding provided both the time and the faculty to conduct this review. The result of over two weeks of very careful review is the ninth version of the Natural World (NW-9) assessment instrument, which provides scores for both Quantitative Reasoning (QR) and Scientific Reasoning (SR). Every item on the test received careful scrutiny. Several items were deleted, most were modified, and several new items were added. We believe this new instrument is much improved. It provides greater balance of items across the objectives and should result in enhanced reliability of both QR and SR scores.

It should also be noted that JMU's General Education Cluster 3 faculty engaged in considerable work on development of faculty expectations for student performances using the Angoff method. Using this methodology, individual faculty members provide a judgment concerning the proportion of students who had completed course requirements successfully that would get individual items correct. Throughout the fall 2007 semester, faculty members were recruited to participate. Thus far, over 40 faculty members that teach in the Cluster 3 area have participated in this process. The Cluster 3 Steering

Committee identified departments that had not participated in the process and made focused inquiries to garner greater support. We continue to collect faculty judgments.

Results in Brief:

- A total of 1,408 entering student tests were randomly assigned to participate in Quantitative Reasoning assessment (<1%) were deleted due to invalid identification or insufficient responses.
- The NW-9 was administered in two formats: paper and pencil (N= 995; 71%) and via computer (N= 413; 29%).
- Overall, the test seems to be functioning quite well. The reliabilities observed for Quantitative Reasoning (α = .64) and the total test, which is how Scientific Reasoning is conceptualized at James Madison (α = .78) are adequate for program evaluation purposes and comparable to those observed for earlier entering first-year cohorts at JMU, which tend to be lower than those obtained with sophomore samples. Though close, the reliability estimates obtained for computer-based administration were slightly lower than those for the paper and pencil format. This will be monitored.
- Entering student performances for QR were at about 62% correct. Cluster Three faculty members are in the process of creating community expectations for student performances through the Angoff procedure. Faculty members will independently provide judgments on how student who had successfully completed cluster requirements would perform on each item. Supplemental reports will compare these ratings with the performances of entering students.
- Students entering the university with Advanced Placement (AP) or Dual Enrollment college credits tended to score higher than students entering without relevant college-level credits.
- Additional analyses produced encouraging results that provide evidence that both AP and Dual Enrollment credits contribute to higher scores on both QR and SR. Correlations for credit hours transferred in and AP credit hours were +.21 and +.28 respectively.

4.2 <u>Note: Post-testing not yet complete:</u> Describe the value-added (or competency) information/data that was collected. This information should include, but is not limited to, quantitative or qualitative summaries of the differences between pre and post assessments or any performance data

Because the Quantitative Reasoning instrument was administered for the first time in fall 2007, we will not have a true repeated measures sample appropriate for value-added reporting until the spring of 2009. We did collect data from a very large set of sophomores in spring 2008. This data was collected on February 12th and will allow a nice cross-sectional value-added comparison to be made. However, we have not analyzed this data as yet.

4.3 <u>Note: Post-testing not yet complete:</u> Describe any additional evidence of value added (or competency); this might include faculty testimony, student retention, or post graduation evidence.

As noted above, we are still in process of gathering faculty judgments via the Angoff method to form a community expectation for student performances. We are particularly pleased with this methodology, because it will allow us to aggregate faculty judgments across items for form expectations by student learning objectives as well as larger competencies, such as Quantitative Reasoning. Our faculty members are very interested in reviewing first-year and sophomore student performances in relation to their newly developed criteria for competency. We look forward to sharing these with the campus community and the State Council in the near future.

Scientific Reasoning

i. Definition of terms

1.1: James Madison University has established five General Education Clusters of learning objectives required for all students. Together, the clusters of JMU's General Education program align with the mission of the University: *"We are a community committed to preparing students to be educated and enlightened citizens who lead productive and meaningful lives."* Cluster Three: The Natural World is comprised of learning objectives most closely associated with Scientific Reasoning. The eight objectives within Cluster Three define scientific reasoning competency for JMU students and faculty; they are listed in section 1.3 below. JMU's conceptualization of scientific reasoning subsumes quantitative reasoning.

1.2: JMU will report the percentage of students meeting faculty expectations for acceptable student performance. The standard setting methodology for the 9th version of the Scientific Reasoning assessment was the Angoff method, and faculty judgments were collected from over 30 faculty members over the 2007-2008 academic year. We now have faculty expectations for student performances at the test and item level.

1.3: Students competent in scientific reasoning will be able to:

- Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudoscience.
- Use theories and models as unifying principles that help us understand natural phenomena and make predictions.
- Recognize the interdependence of applied research, basic research, and technology, and how they affect society.
- Illustrate the interdependence between developments in science and social and ethical issues.
- Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.
- Discriminate between association and causation, and identify the types of evidence used to establish causation.
- Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.

- Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.

These learning objectives provide the framework upon which courses are reviewed, approved, and assessed for the scientific reasoning component of the Natural World.

ii. Methodology

2.1 To demonstrate competency James Madison University will use a pre-post value added analytic approach. JMU's Assessment Day data collection strategy has been in place for over 20 years and provides data on entering students (fall) and those that have completed 45 to 70 credit hours (spring).

2.2 By assessing students as they enter the University and again after they have earned 45 to 70 credits, JMU is in the position to answer questions regarding differences in students as they enter the University (including those with transfer, Advance Placement, dual-enrollment credits). Also, given that some students in the 45 to 70 credit range have completed their scientific reasoning course requirements and others have not, we can compare the performance of students at various stages of completion in an area and examine differences. The assessment design assures that students are assigned to testing rooms on the basis of the last two digits of their student ID numbers; thus, large representative samples are generated. Because the ID numbers do not change, we can assure that students are assigned to take the same instrument again during their sophomore-junior year. This process allows JMU to examine true pre-post differences.

2.3 The 66-item instrument employed is our locally developed Scientific Reasoning instrument (SR-9), which is now in its 9th version. This instrument, which includes our Quantitative Reasoning (QR-9) test, is the focus of a newly funded NSF project, and has shown appropriate psychometric characteristics at JMU and beyond. Cronbach alpha reliability estimates for first-year Fall 2007 students was .78. Review of our Spring 2008 data reveal a Cronbach alpha of .80. The reliability estimates observed are more than adequate for making group-based inferences. Further, our work with earlier versions of the instrument have provided a growing base of validity evidence that support faculty inferences concerning the efficacy of our Natural World instruction and student learning outcomes.

2.4 Two large, representative, and random samples of students are selected for assessment in the Natural World during two Assessment Days: 1) Entering-first-year students during August orientation (approximately 1,000 students); and 2) a sample of students (approximately 900 students) at the midpoint of their undergraduate academic careers [45-70 credit hours] during the annual February Assessment Day. Following this design, a large random, representative sample of students (N = 1,408) entering the University in fall 2007 were administered NW-9.

2.5 As mentioned above, during JMU Assessment Day activities, students are randomly assigned to testing rooms based on the last two digits of the student ID number. Students are then assigned by room to complete cognitive (such as *scientific reasoning*) and non-cognitive (such as *attitudes toward learning)* assessments. This random assignment allows JMU to assess students in general education areas every year. Assessments are administered using standardized testing procedures and highly trained proctors.

2.6 Among the analytical methods used to examine the pre-test scores of entering students on Scientific Reasoning, JMU employs the following:

- Differences—Students who enter the university and are transferring in no credits, vs. those who transfer in AP credits or dual-enrollment credits.
- Competency—The percentage of the students who achieve a faculty developed expectation for student performance. .

2.7 Among the analytical methods used to assess the impact (pre-post) of Cluster Three courses in achieving Scientific Reasoning, JMU employs the following:

- Differences—Students who have completed all Natural World course requirements should perform better than students who have taken only one or two courses related to scientific reasoning.
- Change or Value-Added—Sophomore and junior-level students should perform better on the Scientific Reasoning test than entering first-year students. Further, students who have completed their Natural World course requirements should perform better on the Scientific Reasoning test than they did as entering first-year students [Value-Added].
- Competency—The percentage of the students who have completed their Cluster Three scientific reasoning coursework that achieve the standard or expectation established by JMU faculty. The performance level associated with competency was determined through a modified Angoff standard setting procedure that allows faculty judgments on individual items to be aggregated to form expectations for each learning objective as well as the total test. Over 30 Cluster Three faculty have participated in this process to date.

2.8 Establishing Change or Value-Added:
Because students are assigned to testing locations on the basis of the last two digits of their JMU ID and all entering and sophomore/juniors are required to participate in their respective Assessment Day, JMU can provide both cross-sectional and true repeated-measure samples. Our first true repeated measures sophomore sample using SR-9 will be collected in spring 2009. This design will continue into subsequent years.

The magnitude of proficiency change over time will be calculated by administering the SR-9 at two time points: pretest: August Assessment Day; and posttest: February Assessment Day. The average pretest and posttest scores of students who have completed the relevant general education coursework will be used. This difference score will be standardized by dividing the difference by the pretest standard deviation to enhance interpretation and reporting. Quoting from the *General Education Assessment at JMU Instrument Categories* document [which is available for download at http://www.jmu.edu/assessment/JMUAssess/Gened/GenEd%20Assessment_Instrument%20Categories.pdf ,"The specific standardized mean difference index employed is Cohen's d defined as follows:

$$d = \frac{M_{post} - M_{pre}}{SD_{pre}}$$

with Mpost representing the average posttest raw score, Mpre the average pretest raw score and SDpre the standard deviation of the pretest raw scores. The interpretation of the resulting d is best conveyed using an example: if d = 0.5, the average posttest scores are half a standard deviation unit higher than average pretest scores. In addition to Cohen's d and the descriptive statistics for the pretest and posttest scores, the correlation between pretest and posttest scores should also be reported so that meta-analytic studies that combine the standardized mean difference across cohorts and measures can be pursued." (p. 3). Examinee motivation must be a consideration for appropriate interpretation of performances emanating from testing contexts without personal consequences for individual examinees. Continued inference validation research is being conducted at JMU to address these issues; thus far, our cumulative evidence suggests student motivation is sufficiently high for most students.

iii. Process evaluation

3.1 Annually, reports are generated and distributed to the Cluster Steering Committees of all five clusters, including Cluster Three, the committee assigned to scientific reasoning. In addition, each JMU Cluster is reviewed on a 5-year cycle where program efficacy is carefully monitored and data use for program improvement is reported and monitored.

3.2 The assessment process at JMU is fully integrated with an institutional reporting system by which academic programs complete an annual Assessment Progress Template (APT; see http://www.jmu.edu/assessment/JMUAssess/apt.html) as part of the departmental annual report. All programs also complete Academic Program Review (APR) on a six-year cycle. The newly revised guidelines can be reviewed at http://www.jmu.edu/acadaffairs/wm_library/Revised%20Guidelines%20of%2006-05-07.pdf). Academic programs are expected to report five years of assessment data as part of the APR process. All General Education program clusters are reviewed on a five year cycle, and the General Education program as a whole participates in the University's Academic Program Review process. JMU's 2007 Six Year Plan submitted to SCHEV includes several goals closely associated with scientific reasoning, the continued adoption and full implementation of the APT, strengthening the relationship of the APT to the APR process and decision making (see JMU goals 4.1, 4.2, 4.3, and 4.7). In addition, Goal 4 of JMU's Defining Characteristics for 2006-2012 states the institution's commitment to "Establish and assess optimum competencies for the General Education program" of which scientific reasoning is a member.

3.3 James Madison University's assessment expenditures include funding of the Center for Assessment and Research Studies, as well as providing support across campus for program assessment. Estimated annual cost is $565,755 for the assessment related component of services. It should be noted that this annual cost includes assessment for JMU's five clusters of general education, student affairs assessment and program assessment. Estimated cost for a single cluster, in this case scientific reasoning, is $40,411 annually.

iv. Data presentation

These data are not due until January 2010.

F. Other Significant Accomplishments

1. Faculty
   - ∑ Assembled a group of 13 faculty to develop one or more innovative Cluster Three courses that will have widespread participation by faculty in the cluster, serve the needs of the general non-science major, and be appealing and interesting to these students.
   - ∑ David Pruett and Robert Hanson of the Department of Mathematics and Statistics were awarded a GenEd grant.
   - ∑ Drs. Donna L. Sundre (PI), along with Co-PIs Carol A. Hurney, Christopher G. Murphy, Mary K. Handley, and Richard F. West worked with teams from four other institutions on assessment of scientific and quantitative reasoning. The four institutions are combining their own assessment instruments with the Cluster Three instrument. This work will allow other institutions to adapt JMU's assessment model and JMU to determine the generalizability of the C3 instrument. This work is supported by a grant from the National Science Foundation.

2. Student

   Two students presented at the GenEd Conference on work they completed in Cluster Three courses:

   Ashley Durden, "The Science and Social History of the Harp"

   Tom Fannon, "The Physics of Light and Sound: The Djembe Drum"

3. Other

   The Cluster Coordinator, working with CARS, initiated a research program to determine whether AP exams provide a good measure of mastery of process objectives such as those on the Cluster Three assessment instrument. Ultimately, the information derived from this research should allow us to set cutoff scores for AP credit that reflect mastery of cluster objectives.

G. Statistical Profile

During the 2007-2008 year, Cluster Three enrolled the following numbers of students in each track and group:

Track	Group	Fall 2007	Spring 2008	Total
I	1	3365	2397	5762
	2	2653	1600	4253
	3	2331	2329	4660
	GSCI 104	592	551	1143
II		824	705	1529

The table beginning on the next page shows the number of students enrolled, total spaces made available, and the size of sections offered. Excess spaces offered are indicated in blue, and overrides are indicated in red.

In general, more seats were offered than filled, with the great excess offerings occurring in Track I, Group 3. Relatively large numbers of overrides were given for GSCI 101 and 104, ASTR 120 and 121, and PHYS 140.

	FALL, 2007		SPRING, 2008	
	# Enrolled Students (Total seats available)	Class Size (# sections)	# Enrolled Students (Total seats available)	Class Size (# sections)
Track I				
Group1 GISAT 151	68 (101) +33	24 (2), 19(1), 18(1), 16 (1)	27(26) -1	26(1)
GISAT 251	65 (75) +10	25(3)	5(35) +30	35(1)
MATH 103	416 (417)	32 (13)	308(309)	60(2), 33(3), 30(3)
MATH 205	811 (836) +25	72(2), 71(3), 70(3), 60(2), 59(1), 55(1), 35(1)	697(703) +6	73(1), 70(1), 69(4), 59(1), 54(1), 37(2), 35(1), 32(1), 30(1)
MATH 220	1358(1348) -10	34 (11), 32(12), 31(2), 30 (7), 29(10), 28(1)	1178(1192) +14	34(7), 33(17), 32(4),31(2), 30 (4), 29(2), 25(1)
MATH 220H	20 (20)	20(1)		
MATH 231	339 (350) +11	33(4), 32(5), 30(1), 28(1)	94(99) +5	33(3)
MATH 235	288(309) +21	32(4), 31(1), 30(5)	88(100) +12	25(4)
Group 2 CHEM 120	262(270) +8	135(2)	162(165) +3	165(1)
CHEM 131	603(612) +9	150(4), 12 (1)	306(307) +1	154(1), 153(1)
CHEM 131 LAB	497(552) +55	24(23)	264(264)	24(11)
GISAT 112	230 (238) +9	24 (8), 23 (2)	93(82) - 11	23(2), 18(2)
GSCI 101	484 (446) -38	112(1), 75(1), 70(1), 55(1), 54(1), 40(2)	505(497) -8	100(1), 98(1), 73(1), 48(2), 46(2), 38(1)
GSCI 121	116(120) +4	24(5)	116(115) -1	24(3), 23(1), 20(1)
PHYS 140	186(155) -31	80(1), 50(1), 20(1)	48(35) -13	35(1)
PHYS 140 LAB	171(180) +9	18(10)	64(72) + 8	18(4)
PHYS 215	28(24) -4	24(1)	18(18)	18(1)
PHYS 240	76(70) -6	28(1), 15(2), 12(1)	24(24)	24(1)
Group 3				

ASTR 120	140(100) -40	40(1), 30(2)	0	-
ASTR 121	0	-	112(100) -12	60(1), 40(1)
BIO 114	359(360) +1	24(15)	283(287) +5	24(12)
BIO 114H	17(17)	17(1)	0	-
BIO 270	215(216) +1	24(9)	214(214)	24(9)
GANTH 196	255(231) +6	68(1), 63(1), 40(1), 30(2)	184(174) -10	47(2), 42(2)
GBIO 103	355(359) +4	72(4), 71(1)	260(364) +104	148(1), 147(1), 69(1)
GEOL 110	197(216) +19	24(9)	208(225) +17	25(9)
GEOL 200	0	-	18(18)	18(1)
GEOL 211	55(50) -5	50(1)	158(150) +8	100(1), 50(1)
GGEOL 102	223(224) +1	112(2)	246(310) +64	110(1), 100(2)
GGEOL 115	444(448) +4	112(4)	288(400)+112	100(4)
GISAT 113	71 (72) +1	24 (3)	219(229) +10	30(1), 24(4), 23(1), 22(1), 21(1), 15(1)
GPSYC 122	0	-	139(140) +1	140(1)
Lab **GSCI 104**	592(536) -56	72(1), 30(2), 20(1), 18(5), 17(1), 16(11), 15(2), 12(2), 10(10), 7(1),	551(514) -37	24(3), 20(7), 16(4), 15(8), 14(1), 13(4), 11(2)
Track II **MATH 107**	187 (200) +13	25(8)	104(108) +4	27(4)
GSCI 161	133(134) +1	24(3), 22(2), 18(1)	133(135) +2	30(1), 25(2), 20(2), 15(1)
GSCI 162	134(130) +4	24(2), 22(2), 20(1), 18(1)	139(136) -3	31(1), 25(2), 20(2), 15(1)
GSCI 163	119(136) +17	24(1), 20(3), 18(2), 16(1)	100(100)	20(5)
GSCI 164	123(136) +13	24(1), 20(3), 18(2), 16(1)	101(115) +14	35(1), 20(4)
GSCI 165	128(169) +41	25(1), 24(7)	128(177) +49	35(3), 24(3), 16(1)

Over 80% of the sections in Cluster Three were taught by full-time faculty (Fall 2007: 80%; Spring 2008: 84%). The heaviest reliance on part-time and graduate student instructors occurred in lab courses (CHEM 131L, PHYS 140L, BIO 114, BIO 270).

The number of sections taught by full-time faculty, part-time faculty, and graduate students are given in the table below for each course during each semester of the 2007-2008 academic year.

	FALL, 2007			**SPRING, 2008**		
	# Sections taught by Full-time Faculty*	**# Sections taught by Part-time faculty****	**# Graduate Assistant**	**# Sections taught by Full-time faculty***	**# Sections taught by Part-time faculty****	**# Graduate Assistant**
Track I						
Group1 **GISAT 151**	5	0	0	1	0	0
GISAT 251	3	0	0	1	0	0
MATH 103	12	1	0	8	2	0
MATH 205	13	3	0	8	5	0
MATH 220	36	7	0	38	0	0
MATH 231	12	0	0	-	-	-
MATH 235	10	0	0	4	0	0
Group 2 **CHEM 120**	2	0	0	1	0	0
CHEM 131	5	0	0	2	0	0
CHEM 131 LAB	12	11	0	5	6	0
GISAT 112	11	0	0	4	0	0
GSCI 101	6	1	1	8	0	0
GSCI 121	5	0	0	5	0	0
PHYS 140	2	0	0	1	0	0
PHYS 140 LAB	5	5	0	2	2	0
PHYS 215	1	0	0	1	0	0
PHYS 240	4	0	0	1	0	0
Group 3 **ASTR 120**	3	0	0	-	-	-

ASTR 121	-	-	-	2	0	0
BIO 114	Lecture 3 Lab 10	Lecture 0 Lab 2	Lecture 0 Lab 3	Lecture 4 Lab 6	Lecture 0 Lab 1	Lecture 0 Lab 5
BIO 270	Lecture 2 Lab 1	Lecture 0 Lab 8	Lecture 0 Lab 0	Lecture 2 Lab 1	Lecture 0 Lab 8	Lecture 0 Lab 0
GANTH 196	5	0	0	4	0	0
GBIO 103	3	2	0	2	1	0
GEOL 110	9	0	0	9	0	0
GEOL 200	0	-	-	1	0	0
GEOL 211	1	0	0	2	0	0
GGEOL 102	1	1	0	2	1	0
GGEOL 115	3	1	0	4	0	0
GISAT 113	3	0	0	8	0	0
GPSYC 122	-	-	-	1	0	0
Lab **GSCI 104**	30	6	0	30	3	0
Track II **MATH 107**	8	0	0	4	0	0
GSCI 161	6	0	0	6	0	0
GSCI 162	5	1	0	6	0	0
GSCI 163	5	2	0	3	2	0
GSCI 164	5	2	0	3	2	0
GSCI 165	7	7	0	7	0	0

* Faculty are classified as full-time if they are listed as such in the electronic directory or a departmental home page. Emeritus faculty are classified as full-time.

** Faculty are classified as part-time if the person is listed as part-time faculty in the electronic directory or a departmental home page, or if they are an administrator without a rank of assistant professor or higher. However, some faculty classified as part-time may have extensive teaching experience at JMU.

- Specific resource needs
    - Instructional personnel
        - Full-time
          The greatest area of need currently is faculty to teach GSCI 101. The initiation of the new engineering program has drawn faculty that normally teach in GSCI 101 to courses that support the new program.

        - Part-time
          No needs in this area.

        - Graduate assistant support
          No needs in this area.

    - Technology
      No needs in this area.

    - Library
      No needs in this area.

    - Equipment
      No needs in this area.

    - Classroom Space
      There is a continued need for classrooms with physical space more conducive to the pedagogical needs of the non-major. Especially needed are classrooms that seat 50-75 students with moveable tables, rather than bolted seats in rows, and computer work stations around the perimeter of the room to allow Internet work during class.

H. Cluster Activities planned for next academic year

1. Conduct Cluster Three Self-Study

   The Cluster Coordinator will assemble a committee to complete the self-study to be presented to the GEC either at the end of Spring 2009 or at the beginning of Fall 2009.

2. Fully implement predictive model for determine enrollment needs in Cluster Three.

   During Summer 2008, the Cluster Coordinator will collect data from departments and through queries to build the predictive model. The results of the model will be shared with department to determine whether the model needs to be revised.

3. Update how AP credits are counted toward Cluster Three and send to departments for approval.

   The Cluster Coordinator will propose several changes to how AP credits are counted and will work with departments to implement these changes.

4. Continue work on new assessment goals.

   In early Fall 2008, the Cluster Three Committee will review the Spring 2008 Cluster Three Assessment report and provide the interpretation of the results presented in the report. This interpretative report with be shared with Cluster Three faculty.

   Work will continue on determining which objectives are mastered and which are not, and once this work is completed, the committee will turn to determining the common misconceptions students have with respect to the cluster objectives.

APPENDIX A
Assessment Progress Template
For Annual Academic Department Reporting

Program: Cluster Three of General Education

I. **Objectives** - Please provide your academic program's learning goals and objectives.

The learning objectives for Cluster Three of General Education are as follows:

- Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudoscience.
- Use theories and models as unifying principles that help us understand natural phenomena and make predictions.
- Recognize the interdependence of applied research, basic research, and technology, and how they affect society.
- Illustrate the interdependence between developments in science and social and ethical issues.
- Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.
- Discriminate between association and causation, and identify the types of evidence used to establish causation.
- Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.
- Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.

Describe the process by which these receive faculty review.

The objectives are reviewed periodically by the Cluster Three Committee, which is composed of faculty representatives from each package and departments in the cluster, from the SGA, from CARS, and from Institutional Effectiveness. Review is initiated in response to feedback from committee members, faculty teaching in the cluster, or assessment results.

No objectives were modified, deleted, or added in the last year.

II. **Course/Learning Experiences** - Provide the linkage between your program's goals and objectives and their instructional delivery via your curriculum. [This can be demonstrated with a matrix that lists the goals and objectives by the courses that address each. See the attached form below.]

Mapping program objectives to curriculum

Objectives:	Courses/ Experiences Where Objective is Addressed:
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudoscience.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 161, 162, 163, 164, 165
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 Track II: GSCI 161, 162, 163, 164
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	Track I *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 Track II: GSCI 162, 164, 165
4. Illustrate the interdependence between developments in science and social and ethical issues.	Track I *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 162, 165
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 161, 162, 163, 164, 165

Mapping program objectives to curriculum (cont.)

Objectives:	Courses/ Experiences Where Objective is Addressed:
6. Discriminate between association and causation, and identify the types of evidence used to establish causation.	Track I *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 161, 162, 163, 164, 165
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	Track I *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 Track II: GSCI 161, 163
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 162, 163, 165

III. **Evaluation/Assessment Methods** - Provide a listing of the systematic methods and procedures for gathering information about achievement of your goals and objectives. [This can also be demonstrated with a matrix that lists the goals and objectives by assessment methods. See the attached form below.] Please also describe the process for systematic data collection.

See Assessment Report

IV. **Objective Accomplishments/Results** - Provide a description of your program's assessment results for the last two years—and more if you wish to report them.

See Assessment Report

V. **Dissemination-** Describe how your assessment results are shared with your faculty and others concerned with your program. Illustrate how your assessment results are incorporated in the planning and governance structure of your program.

Results of the assessment are posted on the Cluster Three Blackboard organization and email to the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.

VI. **Uses of Evaluation/Assessment Results and Actions Taken** - In the attached matrix, demonstrate how the program's assessment results have been used to contribute to program improvement and enhanced student learning and growth. The matrix can be used to note assessment findings from one or more sources and how the program faculty responded. Examples of program actions taken might include curriculum revisions, instructional delivery changes, changes in course sequencing, or increased emphasis on specific skill development.

Although Cluster Three assessment pretty clearly demonstrates an effect of cluster courses on mastery of cluster objectives, our assessment instruments have not had sufficient reliability with respect to individual objectives to allow faculty to use information about performance on specific objectives to inform their teaching. In Fall 2007, the Cluster Three Committee will address the question of what direction to take cluster assessment in the future.

Matrix of six key elements in program assessment template

Objectives	Course/ Learning Experiences	Evaluation/ Assessment Methods	Results (% correct)	Dissemination	Uses of Evaluation
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudoscience.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 161, 162, 163, 164, 165	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 Track II: GSCI 161, 162, 163, 164	See Assessment Report:	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).

Matrix of six key elements in program assessment template

3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	Track I *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 Track II: GSCI 162, 164, 165	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).
4. Illustrate the interdependence between developments in science and social and ethical issues.	Track I *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 162, 165	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101,	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of

	121, PHYS 140, 215, 240 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 <u>Track II</u>: GSCI 161, 162, 163, 164, 165			are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	overall assessment score in curricular revision).
6. Discriminate between association and causation, and identify the types of evidence used to establish causation.	<u>Track I</u> *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 <u>Track II</u>: GSCI 161, 162, 163, 164, 165	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	<u>Track I</u> *Group 2*: CHEM 120, 131, GISAT 112, GSCI 101, 121, PHYS 140, 215, 240 <u>Track II</u>: GSCI 161, 163	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment

Matrix of six key elements in program assessment template

				to the Cluster Three Committee, who in turn share these with the faculty they represent.	score in curricular revision).
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	Track I *Group 1*: GISAT 151, 251, MATH 103, 107, 205, 220, 231, 235 *Group 3*: BIO 114, 270, GANTH 196, GBIO 103, GEOL 110, 200, 211, GGEOL 102, GISAT 113, GPSYC 122, GSCI 115, PHYS 120, 121 Track II: GSCI 162, 163, 165	See Assessment Report	See Assessment Report	Assessment results are posted on the Cluster Three Blackboard organization and emailed to the members of the organization. They are also distributed to the Cluster Three Committee, who in turn share these with the faculty they represent.	Reliability of assessment of individual objectives is too low to permit use as basis for action (see narrative for discussion of use of overall assessment score in curricular revision).

APPENDIX B

Cluster Three: The Natural World
Cluster Assessment Results and Interpretation

Fall 2007 Assessment Results

Carol Barry
Ross Markle
Cliff Wigtil
Donna L. Sundre
Center for Assessment and Research Studies
James Madison University
November 12, 2007

Cluster 3 Assessment Report
Fall 2007

Executive Summary

This report provides the first results for the ninth version of the Natural World (NW-9) assessment instrument. The instrument was carefully reviewed during the summer of 2007, and several new items were added. This is the same instrument that will be marketed to many external clients as the Quantitative Reasoning (QR) and Scientific Reasoning (SR) tests. The results obtained with fall 2007 new entering first-year students were overall quite good, and the Cluster 3 Committee hopes to administer this instrument for several years to come.

Results in Brief:

- A total of 1,408 entering student tests were available for this analysis; 13 (<1%) were deleted due to invalid identification or insufficient responses.
- The NW-9 was administered in two formats: paper and pencil (N= 995; 71%) and via computer (N= 413; 29%).
- Overall, the test seems to be functioning quite well. The reliabilities observed for Quantitative Reasoning ($\alpha = .64$) and the total test, which is how Scientific Reasoning is conceptualized at James Madison ($\alpha = .78$) are adequate for program evaluation purposes and comparable to those observed for earlier entering first-year cohorts at JMU, which tend to be lower than those obtained with sophomore samples. Though close, the reliability estimates obtained for computer-based administration were slightly lower than those for the paper and pencil format. This will be monitored.
- Entering student performances for QR were at about 62% correct and 64% correct for the total SR score. Cluster Three faculty members are in the process of creating community expectations for student performances through the Angoff procedure. Faculty members will independently provide judgments on how student who had successfully completed cluster requirements would perform on each item. Supplemental reports will compare these ratings with the performances of entering students.
- Students entering the university of Advanced Placement (AP) or Dual Enrollment college credits tended to score higher than students entering without relevant college-level credits. Increases of 4-10% in QR and SR scores were observed.
- Additional analyses produced encouraging results that provide evidence that both AP and Dual Enrollment credits contribute to higher scores on both QR and SR.

Fall 2007, Cluster 3 Assessment Report Outline

Introduction

This report describes the results of the Fall 2007 Assessment Day administration of the Natural World test version 9 (NW-9). The NW-9 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. Fall 2007 was the first time the NW-9 was administered to first-year JMU students. This report contains the results of statistical analyses that describe the relationship between the CL3 program and NW-9 scores. These analyses lend support to the perception that students are learning as a result of participation in CL3 program.

Sample and Data collection procedures

1,421 students participated in the Cluster 3 assessment during the Fall 2007 assessment day. The responses from 13 students were deleted because these students either omitted 50% or more of the items or did not provide a valid JMU id. The data collected in Fall 2007 is the first cycle of students to complete the Natural world test version 9 (NW-9) as first-year students. These students will complete the test again as students with 40-75 credits. The test was administered in two formats; 995 students completed the test using paper and pencil and 413 students completed the test on a computer.

Assessment Strategies Overview

Assessment results are examined in multiple ways. First, scores on each of the subscales are reported for the sample overall. Scores are then analyzed by the number of cluster, AP and transfer courses taken, the relationship with cluster, AP and transfer/dual enrollment credit hours, and finally the relationship with AP test scores. Currently, scores are not examined in relation to a standard of proficiency because standards for performance on the NW-8 are still being established.

Reliability and Scores

Table 1 on page 4 shows the learning objectives assessed by the NW-9 along with the average number of items correct and percent correct scores for the overall sample. Reliability information (Cronbach's coefficient alpha α) is provided for each objectives scale. Reliability values range from 0 to 1 and indicate the degree to which the scores are free from random error. In general, reliabilities above .70 are considered adequate for program evaluation or research. Scores for scales with reliabilities lower than .60 should be interpreted with caution. Reliability and scale scores are reported separately for computer-based and paper-based administrations in Appendix I. Additionally, a graph and table of the p-values (i.e., or proportion of the sample who answered each item correctly) for all items is presented in Appendix II.

Comparability of scores

Because this was the first administration of the NW-9, raw scores reported in Table 1 are not directly comparable to previous cohorts of first-year students.

Table 1. Nw-9 Scores for Fall 2007 (N = 1,408)

Cluster 3 - Learning Objectives	Item(s) Assessing Objective	
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 9.04 (69.53% correct) SD = 1.92 α = .43
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 3.91 (55.89% correct) SD = 1.36 α = .20
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.09 (58.47% correct) SD = 1.64 α = .47
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 5.50 (61.14% correct) SD = 1.44 α = .25
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 12.53 (59.67% correct) SD = 3.12 α = .58
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 5.37 (53.71% correct) SD = 1.84 α = .45
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 14.20 (67.60% correct) SD = 2.99 α = .59
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 6.99 (53.76% correct) SD = 1.89 α = .32
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 16.09 (61.88% correct) SD = 3.73 α = .64
Total Test	1-66	M = 42.18 (63.91% correct) SD = 7.46 α = .78

I. Average Score by Number of Credit Hours

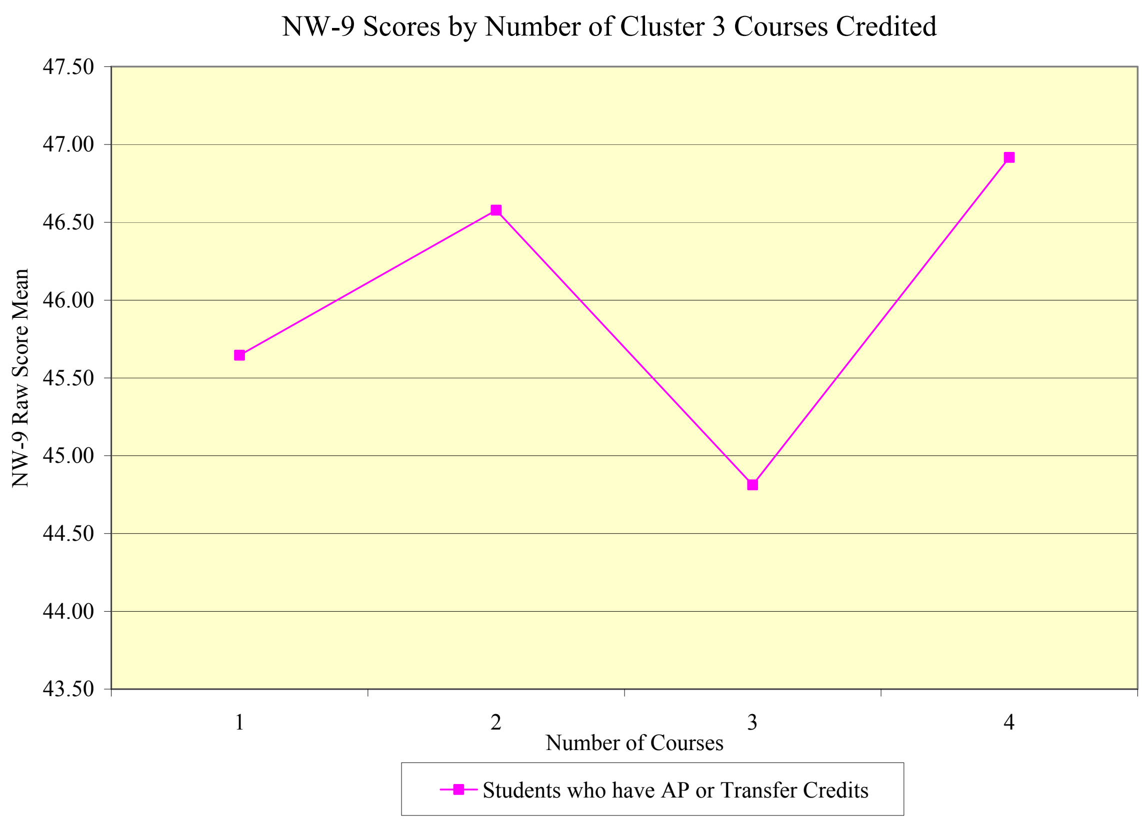
Do students learn or develop more if they have taken more cluster-related courses?

Table 2 presents the average scores by number of credit hours completed. Students who have completed more courses in the cluster should score or perform higher than students who have taken fewer courses in the cluster. Students may have received cluster course credit, through transfer/dual enrollment, AP tests, or even by taking courses at JMU. A visual of these results is presented in Figure 1.

Table 2. Scores by Number of JMU Courses Taken

# of CL 3 credits	N	Raw score Mean	Raw score SD	Percent correct Mean	Percent correct SD
Students who have NO Cluster 3 Credits					
All with no credits	1173	41.45	7.32	62.80	11.09
Students who have AP or Transfer Credits					
1	161	45.65	7.18	69.16	10.87
2	45	46.58	7.49	70.57	11.35
3	16	44.81	6.87	67.90	10.41
4 or more	12	46.92	5.99	71.09	9.08
All students with Credits in Sample					
1	162	45.62	7.16	69.13	10.85
2	45	46.58	7.49	70.57	11.35
3	16	44.81	6.87	67.90	10.41
4 or more	12	46.92	5.99	71.09	9.08

Figure 1.



II. Relationship between NW-9 Scores and Number of Credit Hours Completed

What is the relationship between number of cluster 3 credit hours completed and outcomes?
Correlations between NW-9 scores and number of credit hours completed are presented in Table 3. A positive relationship should exist between number of cluster 3 hours completed and NW-9 scores if the content covered in the courses matches that covered on the NW test.

Correlations can range from -1 to 1. Correlations close to 0 indicate no relationship, while correlations closer to 1 indicate a very high relationship between test scores and credit hours (negative numbers would indicate those who scored high on the test had fewer credit hours).

Table 3. Correlations between Number of Credit Hours Completed and NW-9 Raw Scores

Credit hours	N	NW correlation	NW r^2	QR correlation	QR r^2
No JMU Credit Hours					
Transfer Credit Hours	85	0.21	0.04	0.21	0.04
AP Credit Hours	152	0.26	0.07	0.28	0.08

The graphs below illustrate instability in the correlations due to sampling error (Figures 2 and 3). The error bands show the 95% confidence band around the correlation estimates. Correlations with overlapping bands are not statistically different (at $p < .05$).

Figure 2.

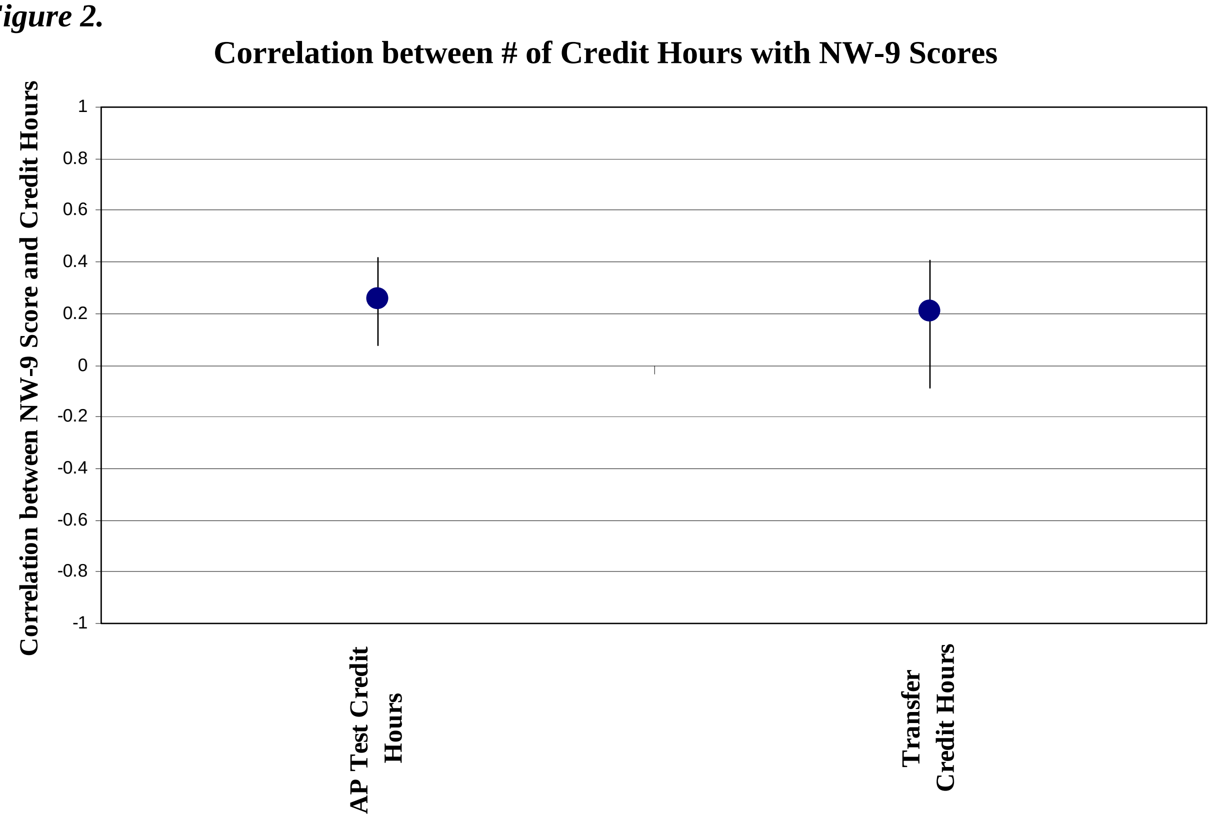
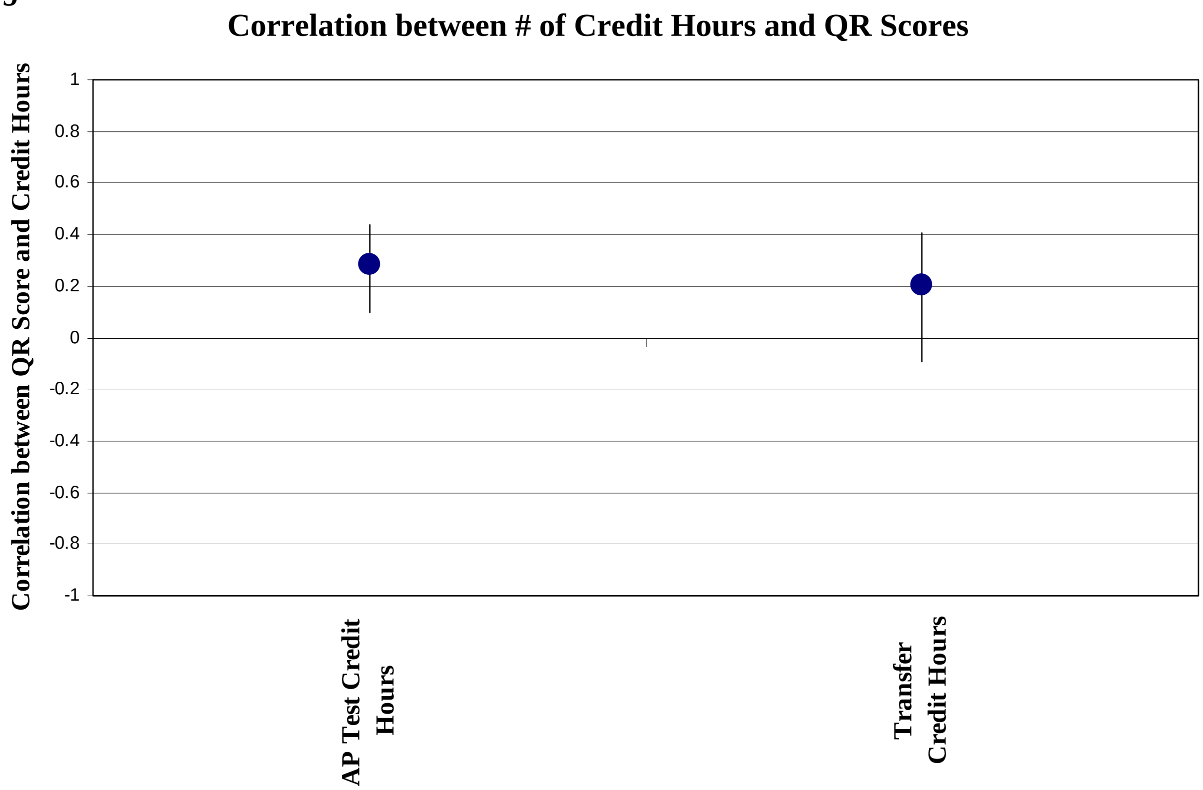


Figure 3.


Correlation between # of Credit Hours and QR Scores
Correlation between QR Score and Credit Hours
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
AP Test Credit Hours
Transfer Credit Hours

III. Relationship between NW-9 Scores and AP Test Score

***What is the relationship between average AP test scores and outcomes?* Correlations between NW-9 scores and AP test scores are reported in Table 4. Given that students are given cluster course credit if they obtain a certain score or higher on the AP test, a positive relationship should exist if higher scores on AP tests indicate higher levels of mastery of the material. In this situation, AP test scores serve as a proxy for grades.**

Again, correlations can range from -1 to 1. Correlations close to 0 indicate no relationship, while correlations closer to 1 indicate a very high relationship between test scores and AP test scores (negative numbers would indicate those who scored high on the test had lower AP test scores).

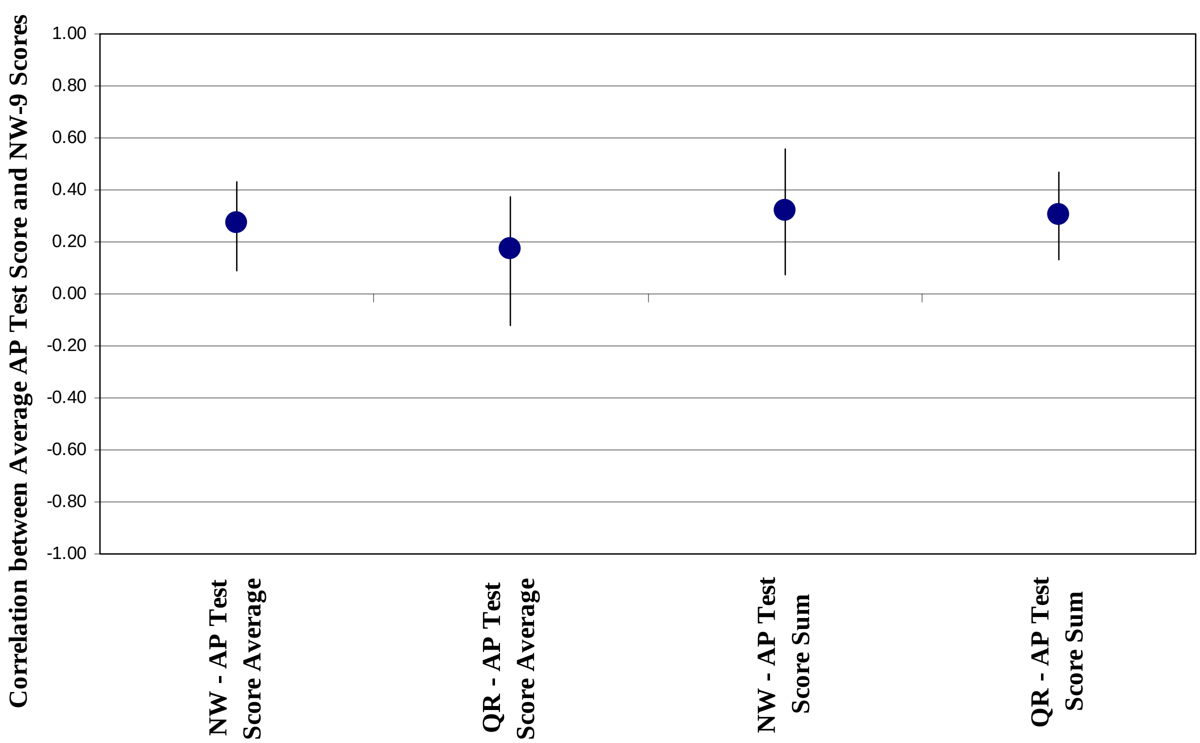
Table 4. Correlations between Average AP test scores and NW-9 Raw Scores

	N	NW correlation	NW r^2	QR correlation	QR r^2
AP Test Score - Average	152	0.27	0.07	0.17	0.03
AP Test Score - Sum	152	0.32	0.10	0.30	0.09

The Figure 4 below illustrates instability in the correlations due to sampling error. The error bands show the 95% confidence band around the correlation estimates. Correlations with overlapping bands are not statistically different (at $p < .05$).

Figure 4.

Correlation between AP Test scores and NW and QR Scores



IV. Meeting a Standard

Do students meet faculty expectations? Students who have cluster credits when they arrive at JMU should be closer to meeting the competency or academic standard than those with no cluster credits.

This assessment strategy does not currently apply to Cluster 3 because a standard for the NW-9 has yet to be established.

Appendix I

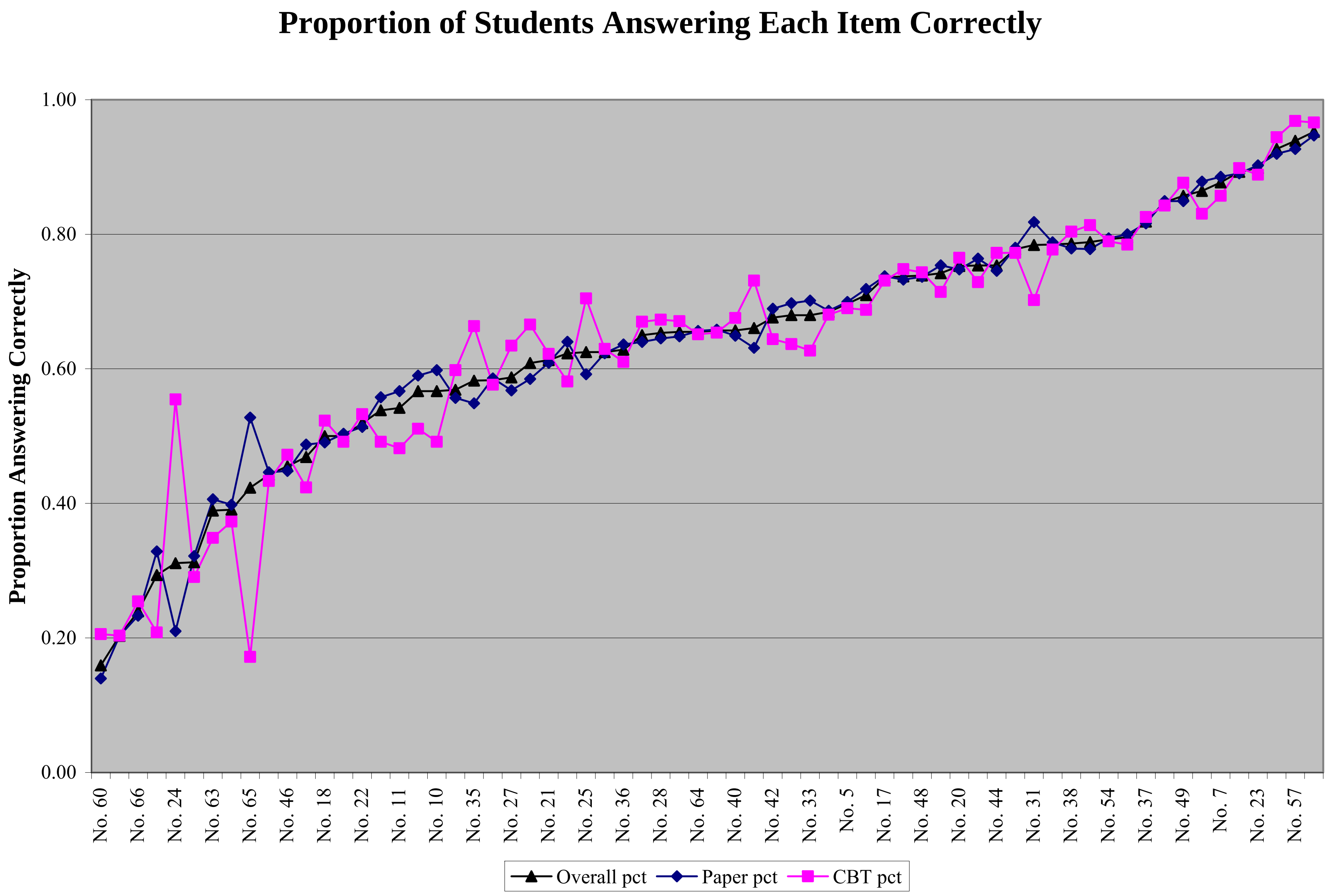
Fall 2007 paper-based Natural World 9 assessment results (N = 995)

Cluster 3 - Learning Objectives	Item(s) Assessing Objective	
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 9.01 (69.34% correct) SD = 1.95 α = .45
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 3.98 (56.91% correct) SD = 1.39 α = .25
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.10 (58.52% correct) SD = 1.68 α = .51
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 5.37 (59.66% correct) SD = 1.43 α = .25
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 12.72 (60.59% correct) SD = 3.18 α = .60
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 5.37 (53.70% correct) SD = 1.84 α = .45
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 14.25 (67.84% correct) SD = 3.09 α = .62
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 6.81 (52.39% correct) SD = 1.83 α = .28
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 16.27 (62.58% correct) SD = 3.78 α = .65
Total Test	1-66	M = 42.29 (64.07% correct) SD = 7.74 α = .79

Fall 2007 computer-based Natural World 9 assessment results (N = 413)

Cluster 3 - Learning Objectives	Item(s) Assessing Objective	
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 9.10 (69.99% correct) SD = 1.84 α = .39
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 3.74 (53.44% correct) SD = 1.26 α = .12
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.08 (58.35% correct) SD = 1.54 α = .39
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 5.83 (64.73% correct) SD = 1.41 α = .22
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 12.07 (57.47% correct) SD = 2.94 α = .51
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 5.38 (53.75% correct) SD = 1.84 α = .47
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 14.08 (67.02% correct) SD = 2.75 α = .50
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 7.42 (57.07% correct) SD = 1.95 α = .38
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 15.66 (60.23% correct) SD = 3.60 α = .60
Total Test	1-66	M = 41.92 (63.51% correct) SD = 6.76 α = .73

Appendix II


Proportion of Students Answering Each Item Correctly
Proportion Answering Correctly
1.00
0.80
0.60
0.40
0.20
0.00
No. 60
No. 66
No. 24
No. 63
No. 65
No. 46
No. 18
No. 22
No. 11
No. 10
No. 35
No. 27
No. 21
No. 25
No. 36
No. 28
No. 64
No. 40
No. 42
No. 33
No. 5
No. 17
No. 48
No. 20
No. 44
No. 31
No. 38
No. 54
No. 37
No. 49
No. 7
No. 23
No. 57
Overall pct
Paper pct
CBT pct

Proportion of Students Answering Each Item Correct (Ascending Order)

Item no.:	Computer-Based pct	Paper pct	Overall pct
No. 60	0.21	0.14	0.16
No. 2	0.20	0.20	0.20
No. 66	0.25	0.23	0.24
No. 53	0.21	0.33	0.29
No. 24	0.55	0.21	0.31
No. 61	0.29	0.32	0.31
No. 63	0.35	0.41	0.39
No. 19	0.37	0.40	0.39
No. 65	0.17	0.53	0.42
No. 43	0.43	0.45	0.44
No. 46	0.47	0.45	0.46
No. 15	0.42	0.49	0.47
No. 18	0.52	0.49	0.50
No. 39	0.49	0.50	0.50
No. 22	0.53	0.51	0.52
No. 32	0.49	0.56	0.54
No. 11	0.48	0.57	0.54
No. 4	0.51	0.59	0.57
No. 10	0.49	0.60	0.57
No. 58	0.60	0.56	0.57
No. 35	0.66	0.55	0.58
No. 6	0.58	0.59	0.58
No. 27	0.63	0.57	0.59
No. 1	0.67	0.58	0.61
No. 21	0.62	0.61	0.61
No. 41	0.58	0.64	0.62
No. 25	0.70	0.59	0.63
No. 51	0.63	0.62	0.63
No. 36	0.61	0.64	0.63
No. 13	0.67	0.64	0.65
No. 28	0.67	0.65	0.65
No. 26	0.67	0.65	0.65
No. 64	0.65	0.66	0.65

Item no.:	Computer-Based pct	Paper pct	Overall pct
No. 30	0.65	0.66	0.66
No. 40	0.68	0.65	0.66
No. 62	0.73	0.63	0.66
No. 42	0.64	0.69	0.68
No. 16	0.64	0.70	0.68
No. 33	0.63	0.70	0.68
No. 45	0.68	0.69	0.68
No. 5	0.69	0.70	0.70
No. 52	0.69	0.72	0.71
No. 17	0.73	0.74	0.74
No. 55	0.75	0.73	0.74
No. 48	0.74	0.74	0.74
No. 3	0.71	0.75	0.74
No. 20	0.77	0.75	0.75
No. 12	0.73	0.76	0.75
No. 44	0.77	0.75	0.75
No. 29	0.77	0.78	0.78
No. 31	0.70	0.82	0.78
No. 34	0.78	0.79	0.78
No. 38	0.80	0.78	0.79
No. 59	0.81	0.78	0.79
No. 54	0.79	0.79	0.79
No. 8	0.78	0.80	0.80
No. 37	0.83	0.82	0.82
No. 50	0.84	0.85	0.85
No. 49	0.88	0.85	0.86
No. 56	0.83	0.88	0.86
No. 7	0.86	0.89	0.88
No. 47	0.90	0.89	0.89
No. 23	0.89	0.90	0.90
No. 9	0.94	0.92	0.93
No. 57	0.97	0.93	0.94
No. 14	0.97	0.95	0.95

APPENDIX C

Cluster 3 Assessment Report

The Natural World

Spring 2008 Sophomore Cohort

THE CENTER FOR ASSESSMENT & RESEARCH STUDIES
MSC 6806
Harrisonburg, VA 22807
540.568.6706 Phone
540.568.7878 Fax
assessment@jmu.edu
www.jmu.edu/assessment

Carol Barry
Amy Thelk
Donna Sundre
May, 2008

Table of Contents

Cluster 3 / Individuals in the Human Community Results from Assessment Day, Spring 2008

Cluster 3 / The Natural World
Spring 2008 Assessment Day

Introduction

This report describes the results of the Spring 2008 Assessment Day administration of the Natural World test version 9 (NW-9). The NW-9 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 2008 was the first time the NW-9 was administered to second-year JMU students. This report contains the results of statistical analyses that describe the relationship between the CL3 program and NW-9 scores. These analyses lend support to the perception that students are learning as a result of participation in CL3 program.

Sample and Data Collection Procedures

One-thousand sixty-seven students participated in the Cluster 3 assessment during the Spring 2008 assessment day. The responses from 15 students were deleted because these students either omitted 50% or more of the items or did not provide a valid JMU ID. Additionally, 32 students were removed from the dataset because they showed evidence of low test-taking motivation. The data collected in Spring 2008 is the first set of students to complete the Natural World test version 9 (NW-9) as students with 40-75 credits. The test was administered in two formats; 923 students completed the test using paper and pencil and 97 students completed the test on a computer.

Assessment Strategies Overview

Assessment results are examined in multiple ways. First, scores on each of the subscales are reported for the sample overall. Scores are then analyzed by the number of cluster, AP and transfer courses taken, the relationship with cluster, AP and transfer/dual enrollment credit hours, and finally the relationship with AP test scores. Currently, scores are not examined in relation to a standard of proficiency because standards for performance on the NW-9 are still being established.

Reliabilities and Scores

Table 1 on the following page shows the learning objectives assessed by the NW-9 along with the average number of items correct and percent correct scores for the overall sample. Reliability information (Cronbach's coefficient alpha α) is provided for each objectives scale. Reliability values range from 0 to 1 and indicate the degree to which the scores are free from random error. In general, reliabilities above .70 are considered adequate for program evaluation or research. Scores for scales with reliabilities lower than .60 should be interpreted with caution. Reliability and scale scores are reported separately for computer-based and paper-based administrations in Appendix I. Additionally, a graph and table of the p-values (i.e., or proportion of the sample who answered each item correctly) for all items is presented in Appendix II.

Comparability of Scores

Because this was the first administration of the NW-9 to students with 45-70 credit hours, raw scores reported in Table 1 are not directly comparable to previous cohorts of second-year students, nor can pre- post-comparisons be made.

Cluster 3 - Learning Objectives	Item(s) Assessing Objective

Table 1. NW-9 Scores for Spring 2008 (N = 1,020)

1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 10.00 (76.89% correct) SD = 1.81 α = .46
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 4.70 (67.16% correct) SD = 1.41 α = .32
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.46 (63.74% correct) SD = 1.60 α = .48
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 6.52 (72.40% correct) SD = 1.36 α = .28
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 13.80 (65.72% correct) SD = 3.14 α = .61
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 5.88 (58.76% correct) SD = 1.81 α = .43
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 15.40 (73.35% correct) SD = 3.05 α = .65
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 8.38 (64.48% correct) SD = 1.80 α = .33
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 17.55 (67.50% correct) SD = 3.74 α = .66
Total Test	1-66	M = 46.96 (71.16% correct) SD = 7.83 α = .82

Average scores by number of courses

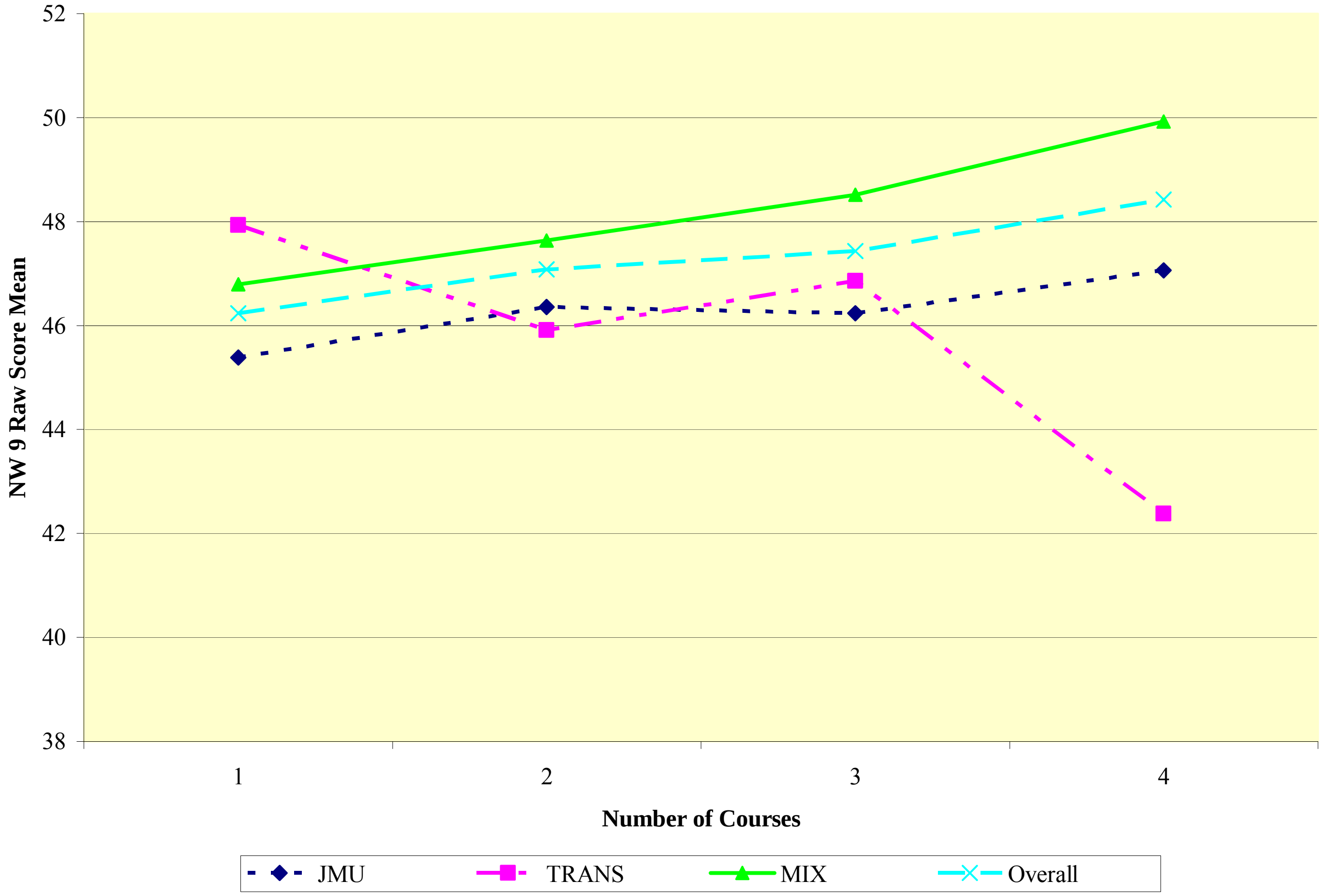
Do students learn or develop more if they have taken more cluster-related courses?

Students who have completed more courses in the cluster should score or perform higher than students who have taken fewer courses in the cluster. A visual of these results is presented in Figure 1.

Table 2. Scores by Number of Cluster 3 Courses Taken

# of CL 3 credits	N	Raw score Mean	Raw score SD	Percent correct Mean	Percent correct SD
Students who have NO Cluster 3 Credits					
All with 0 credit	30	43.53	8.95	65.96	13.56
0 credits, enrolled	20	42.45	9.70	64.32	14.69
0 credits, not enrolled	10	45.70	7.18	69.24	10.88
Students with only JMU Course Credits					
1	119	45.39	8.47	68.77	12.83
2	136	46.35	7.66	70.23	11.61
3	105	46.24	7.22	70.06	10.94
4 or more	35	47.06	10.72	71.30	16.24
Students with only AP Course Credits					
1	2	54.00	2.83	81.82	4.29
2	2	52.50	6.36	79.55	9.64
Students with only Transfer Course Credits					
1	14	47.93	6.23	72.62	9.44
2	11	45.91	4.53	69.56	6.86
3	7	46.86	8.15	71.00	12.35
4 or more	13	42.38	9.27	64.22	14.05
Students with a Mix of Course Credits					
1	110	46.79	7.29	70.90	11.04
2	183	47.63	7.54	72.17	11.43
3	119	48.51	7.19	73.50	10.89
4 or more	84	49.93	7.95	75.65	12.04
All students with Credits in Sample					
1	245	46.23	7.85	70.05	11.90
2	332	47.08	7.52	71.34	11.40
3	231	47.43	7.29	71.86	11.04
4 or more	132	48.42	9.13	73.37	13.83

Figure 1.
NW-9 Scores by Number of Cluster 3 Courses Credited



Relationship between NW-9 scores and number of credit hours completed

What is the relationship between NW-9 scores and number of Cluster 3 credit hours completed?

Correlations between NW-9 scores and number of credit hours completed are presented in Table 3. A positive relationship should exist between number of cluster 3 hours completed and NW-9 scores if the content covered in the courses matches that covered on the NW test.

Correlations can range from -1 to 1. Correlations close to 0 indicate no relationship, while correlations closer to 1 indicate a very high relationship between test scores and credit hours (negative numbers would indicate those who scored high on the test had fewer credit hours).

Table 3. Correlations between Number of Credit Hours Completed and NW-9 Raw Scores

Credit hours	N	NW correlation	NW R^2	QR correlation	QR R^2
Credit Hours Obtained from a Single Source					
JMU credits only	398	0.06	0.00	0.04	0.00
AP credits only	4	--	--	--	--
Transfer credits only	45	-0.27	0.07	-0.14	0.02
Credit Hours Obtained for a Mix Of Sources					
JMU credit hours	451	0.11	0.01	0.14	0.02
AP credits hours	103	0.16	0.03	0.23	0.05
Transfer credit hours	171	0.01	0.00	-0.04	0.00
Combined credit hours	496	0.14	0.02	0.15	0.02
Overall Sample					
Combined credit hours	943	0.09	0.01	0.10	0.01

The graphs below illustrate instability in the correlations due to sampling error (Figures 2, 3, and 4). The error bands show the 95% confidence band around the correlation estimates. Correlations with overlapping bands are not statistically different (at $p < .05$).

Figure 2.
Correlation between # of Credit Hours with NW Scores – Students with only JMU or Transfer Credits

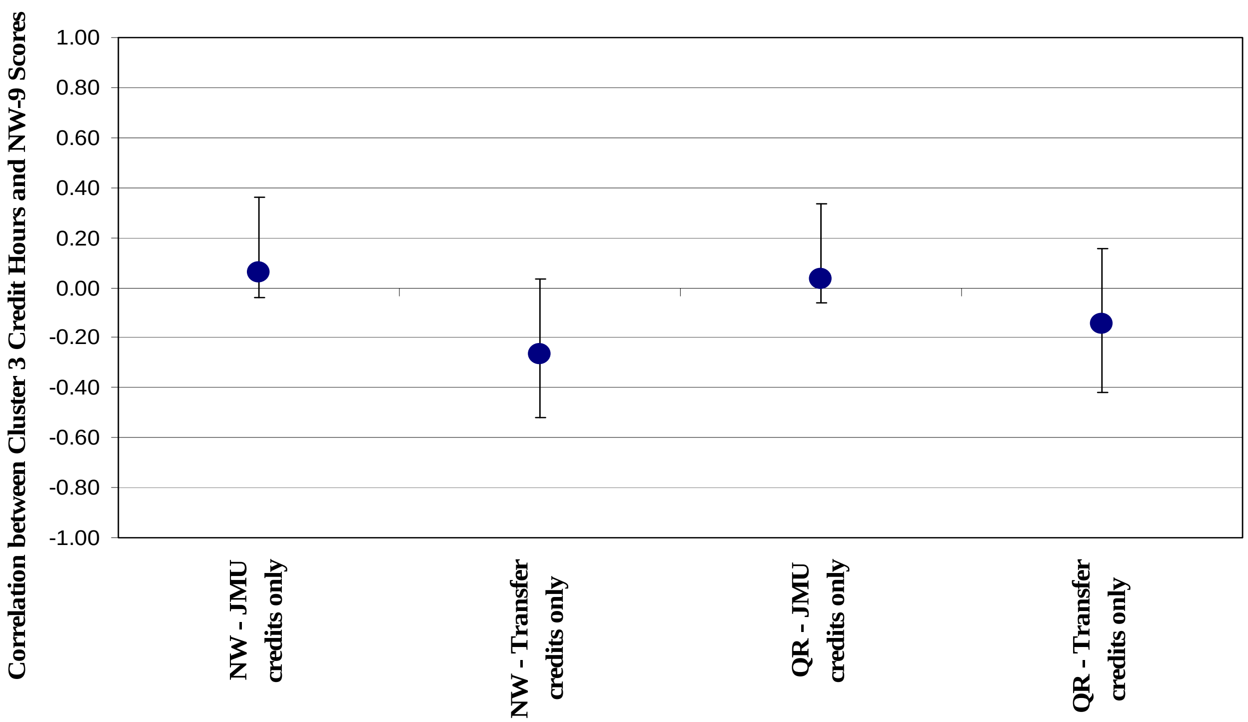


Figure 3.
Correlation between # of Credit Hours with NW Scores – Students with a Mix of Credits

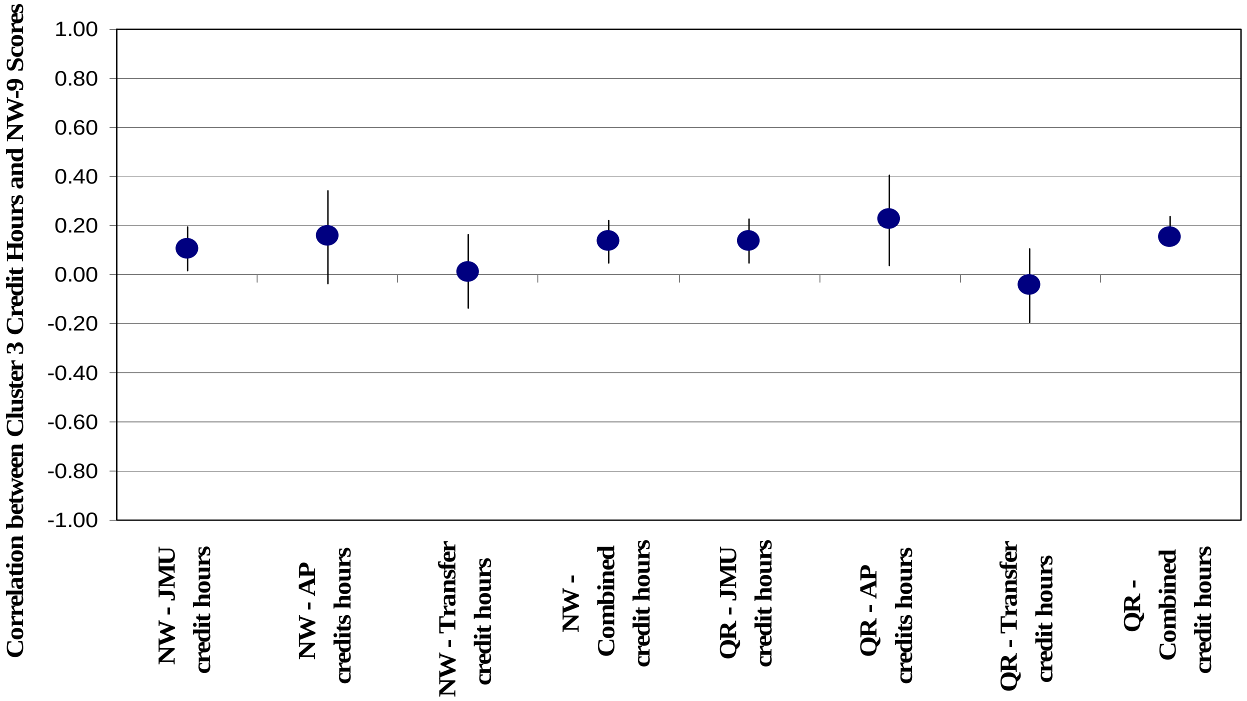
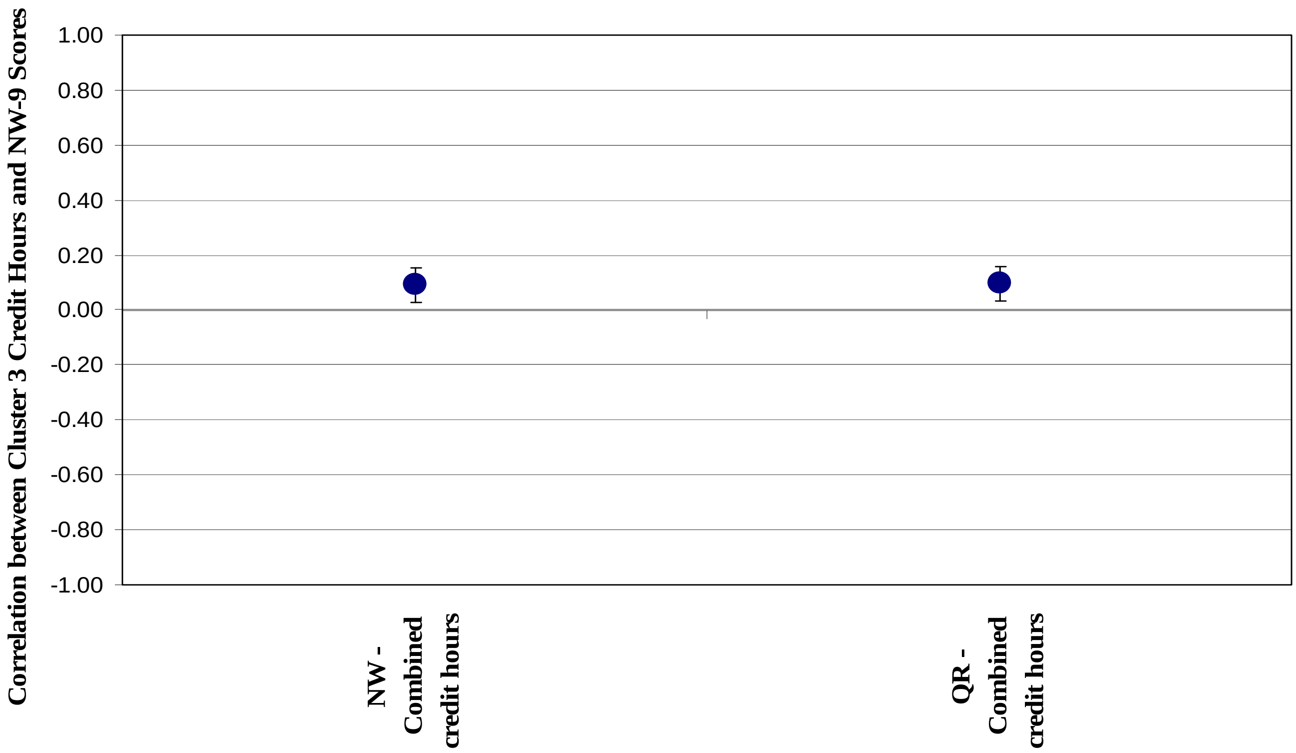


Figure 4.
Correlation between # of Credit Hours with NW Scores – Overall Sample



Relationship between NW-9 Scores and AP Test Score

What is the relationship between average AP test scores and outcomes?

Correlations between NW-9 scores and AP test scores are reported in Table 4. Given that students are given cluster course credit if they obtain a certain score or higher on the AP test, a positive relationship should exist if higher scores on AP tests indicate higher levels of mastery of the material. In this situation, AP test scores serve as a proxy for grades.

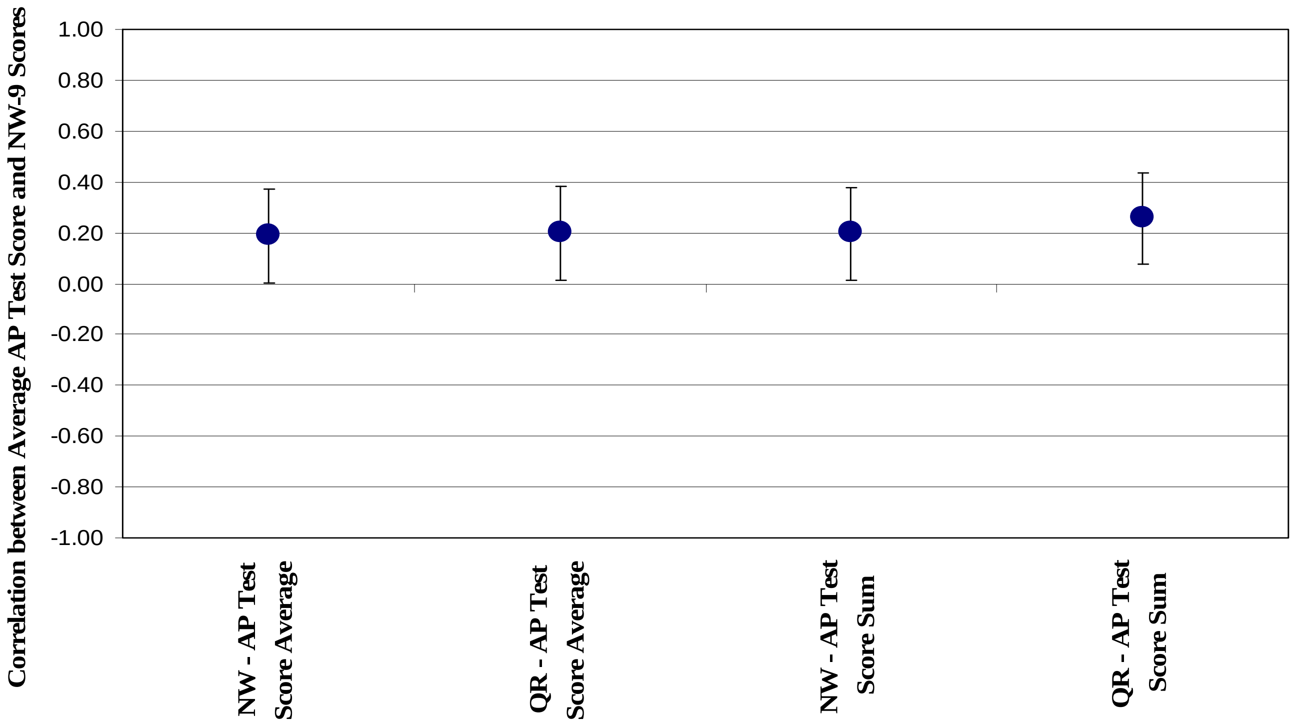
Again, correlations can range from -1 to 1. Correlations close to 0 indicate no relationship, while correlations closer to 1 indicate a very high relationship between test scores and AP test scores (negative numbers would indicate those who scored high on the test had lower AP test scores).

Table 4. Correlations between Average AP test scores and NW-9 Raw Scores

	N	NW correlation	NW R^2	QR correlation	QR R^2
Students with only AP Course Credits					
AP Test Score - Average	4	--	--	--	--
AP Test Score - Sum	4	--	--	--	--
Students with a Mix Of Course Credits					
AP Test Score - Average	103	0.20	0.04	0.20	0.04
AP Test Score - Sum	105	0.20	0.04	0.26	0.07

Figure 5 below illustrates instability in the correlations due to sampling error. The error bands show the 95% confidence band around the correlation estimates. Correlations with overlapping bands are not statistically different (at $p < .05$).

Figure 5.
Correlation between AP Test scores and NW and QR Scores



Meeting a Standard

The Standard Setting Process

For past versions of the Natural World exam, attempts have been made to set standards to facilitate later interpretation of scores. Specifically, faculty members and other users of the data and results tend to express confusion about what the achieved scores on the exam mean: are they "good enough," or are students failing to achieve the learning objectives of the general education program?

In the summer of 2007, a standard setting workshop was conducted on the NW-8, using the Bookmark method. This technique incorporates empirical data to order the items from least to most difficult. A group of content experts (in this case, Cluster 3 faculty) were then asked to find the location in the stack of ordered items that represent the "cut score" – the score that separates the students who are minimally competent in quantitative and scientific reasoning from those who are competent. Obviously, the resulting order of difficulty depends on the students in the sample, and the faculty members involved in this process were not confident setting a standard in this manner.

Early in the Spring 2008 semester, select Cluster 3 faculty participated in an Angoff standard setting. This process involves asking judgments about the likelihood that a "minimally competent" student could get the item correct; a judgment is made for each item. An advantage of this rating process can be completed prior to any data collection, and the results will not capitalize on any idiosyncrasies of a given sample. The ratings for each item (expressed as proportions are then summed over the entire test and the resulting total is the cut score for the exam. Each judge's score was computed, the median of the scores is used as the cut sore. A total of 37 Cluster 3 faculty participated in the Angoff standard setting.

Cut scores, or standards, were computed for each objective for the exam, as well as for the total test (SR) and for the quantitative reasoning (QR) subscale.

Do students meet faculty expectations?

Students who have cluster credits when they arrive at JMU should be closer to meeting the competency or academic standard than those with no cluster credits.

In the table below, the first two columns contain information about the objectives and subscores. The first shaded column contains the cut score, or faculty expectation, derived via Angoff standard setting process, fort the corresponding objective or scale to its left. The top figure is the raw score that is the cut score, and the bottom number is the percentage correct on the objective that must be obtained in order to meet the standard. The remaining columns display the mean score for the group described in the column heading, as well as the percentage of the group meeting the faculty expectation.

The table shows a general trend that, as students are exposed to Cluster 3 coursework, the performance on the exam increases. The lowest means can be seen in the column for first-year students with no Cluster 3 coursework. As the number of courses increases, the means and percentage of students achieving the standard increases, to varying degrees.

Table 5. Fall 2007 First-Year Student and Spring 2008 Sophomore/Junior Performance Compared with Faculty Expectations by Learning Objective

Mean
Percent meeting standard

Objective	# items	Faculty Standard* (% of obj)	Freshmen with no CL3 coursework *n*=1173	Students with 45-70 credits (Spring sample): All Spring 2008 examinees *n*=973	No CL3 coursework *n*=10	Package completers *n*=156
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	13	10.4 (80%)	8.9 20.3%	10.0 44.3%	9.1 20.0%	10.3 50.0%
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	7	5.1 (73%)	4.2 18.4%	4.7 33.1%	4.8 20.0%	4.8 36.5%
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	7	5.3 (76%)	4.0 18.7%	4.5 27.5%	4.1 2.0%	4.7 34.0%
4. Illustrate the interdependence between developments in science and social and ethical issues.	9	7.1 (79%)	6.0 14.2%	6.5 24.5%	6.2 10.0%	6.7 25.6%
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	21	15.7 (75%)	12.3 14.7%	13.8 31.6%	14.0 30.0%	14.3 44.9%
6. Discriminate between association and causation, and identify the types of evidence used to establish causation.	10	7.6 (76%)	5.3 11.7%	5.9 19.6%	5.9 20.0%	6.2 26.3%
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	21	16.3 (78%)	13.9 19.9%	15.4 40.5%	14.6 30.0%	15.7 48.7%
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	13	9.8 (75%)	7.5 14.2%	8.4 27.1%	8.0 20.0%	8.6 34.0%
QR-9	26	19.4 (75%)	15.8 15.4%	17.6 32.2%	18.0 50.0%	18.2 43.6%
NW-9 Total	66	50.4 (76%)	42.4 13.5%	47.0 36.5%	45.7 30.0%	48.3 47.4%

Figure 6.
Percent of students with varying Cluster 3 experience meeting standard on NW-9 objective

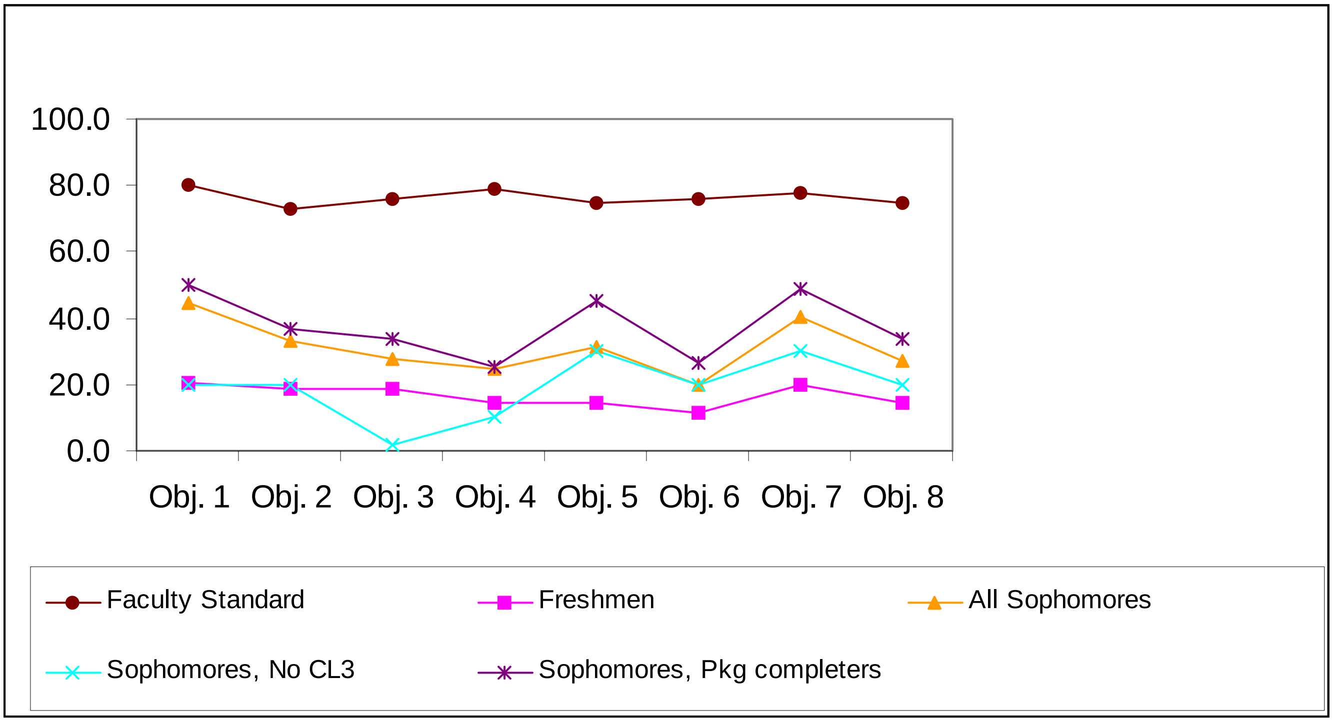
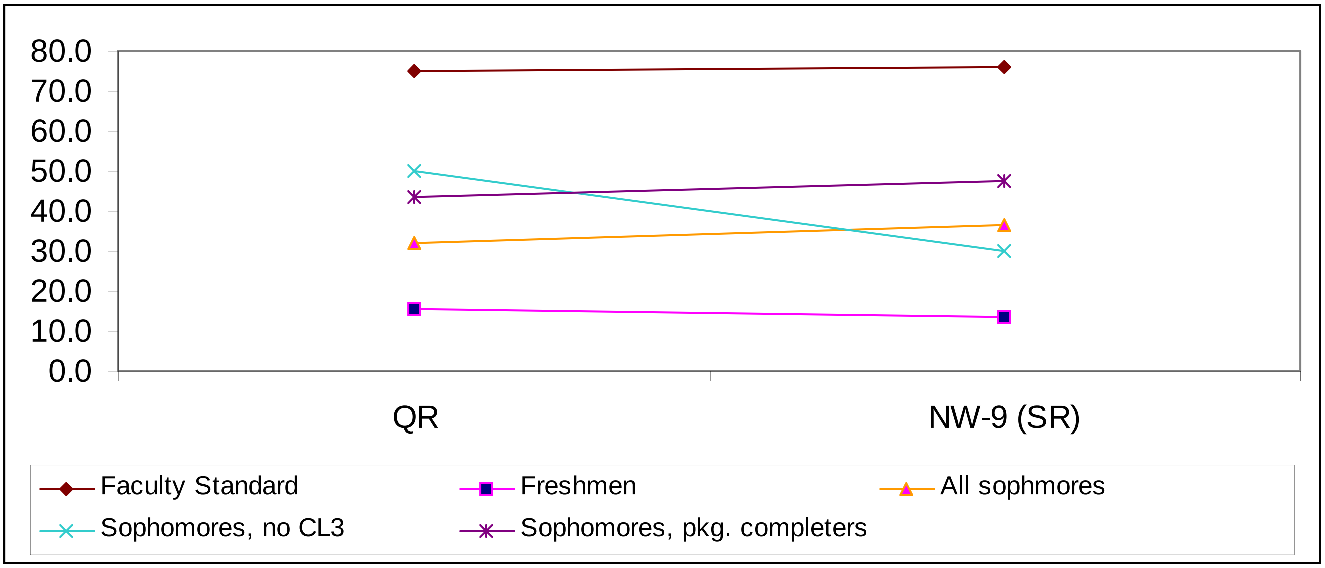


Figure 7.
Percent of students with varying Cluster 3 experience meeting standard on QR and SR scores



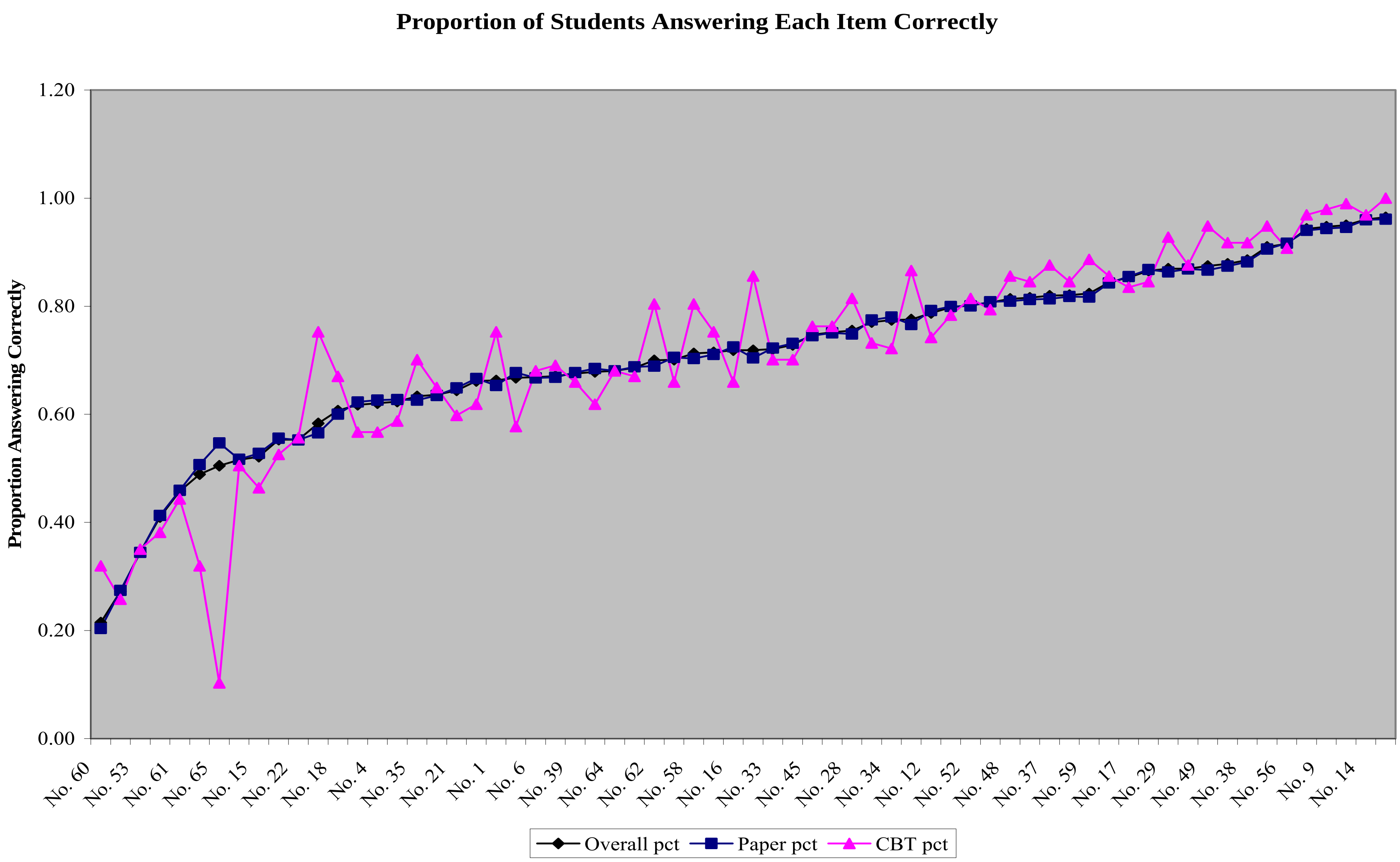
Appendix

Spring 2008 paper-based Natural World 9 assessment results (*N* = 923)

Cluster 3 - Learning Objectives	Item(s) Assessing Objective	
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 9.97 (76.70% correct) SD = 1.84 α = .47
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 4.74 (67.76% correct) SD = 1.43 α = .34
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.48 (64.05% correct) SD = 1.62 α = .49
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 6.49 (72.13% correct) SD = 1.37 α = .30
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 13.80 (65.71% correct) SD = 3.17 α = .62
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 5.85 (58.53% correct) SD = 1.82 α = .43
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 15.35 (73.09% correct) SD = 3.10 α = .66
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 8.33 (64.11% correct) SD = 1.81 α = .34
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 17.54 (67.47% correct) SD = 3.78 α = .67
Total Test	1-66	M = 46.94 (71.12% correct) SD = 8.00 α = .83

Spring 2008 computer-based Natural World 9 assessment results (N = 97)

Cluster 3 - Learning Objectives	Item(s) Assessing Objective	
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	2, 5, 9, 14, 18, 28, 38-41, 55-57 (13 items; 19.7% of test)	M = 10.23 (78.67% correct) SD = 1.48 α = .22
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	17, 20, 22, 27, 64-66 (7 items; 10.6% of test)	M = 4.30 (61.41% correct) SD = 1.23 α = .18
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	1, 15, 16, 43-46 (7 items; 10.6% of test)	M = 4.26 (60.82% correct) SD = 1.43 α = .32
4. Illustrate the interdependence between developments in science and social and ethical issues.	2, 19, 24-26, 29, 55-57 (9 items; 13.6% of test)	M = 6.74 (74.92% correct) SD = 1.17 α = .11
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (21 items; 31.8% of test)	M = 13.82 (65.83% correct) SD = 2.87 α = .52
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	3, 34-37, 53, 60-63 (10 items; 15.2% of test)	M = 6.10 (61.03% correct) SD = 1.76 α = .40
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (21 items; 31.8% of test)	M = 15.91 (75.75% correct) SD = 2.42 α = .47
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	2, 14, 24-26, 29, 38-40, 60-63 (13 items; 19.7% of test)	M = 8.85 (68.04% correct) SD = 1.60 α = .21
Quantitative Reasoning	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63 (26 items; 39.4% of test)	M = 17.63 (67.80% correct) SD = 3.32 α = .57
Total Test	1-66	M = 47.20 (71.51% correct) SD = 5.92 α = .69


Proportion of Students Answering Each Item Correctly
Proportion Answering Correctly
1.20
1.00
0.80
0.60
0.40
0.20
0.00
No. 60
No. 53
No. 61
No. 65
No. 15
No. 22
No. 18
No. 4
No. 35
No. 21
No. 1
No. 6
No. 39
No. 64
No. 62
No. 58
No. 16
No. 33
No. 45
No. 28
No. 34
No. 12
No. 52
No. 48
No. 37
No. 59
No. 17
No. 29
No. 49
No. 38
No. 56
No. 9
No. 14
Overall pct
Paper pct
CBT pct

Proportion of Students Answering Each Item Correct (Ascending Order)

Item no.	Overall pct	Paper pct	Computer-based pct	Item no.	Overall pct	Paper pct	Computer-based pct
No. 60	0.21	0.20	0.32	No. 51	0.72	0.70	0.86
No. 2	0.27	0.27	0.26	No. 33	0.72	0.72	0.70
No. 53	0.35	0.34	0.35	No. 41	0.73	0.73	0.70
No. 63	0.41	0.41	0.38	No. 45	0.75	0.75	0.76
No. 61	0.46	0.46	0.44	No. 40	0.75	0.75	0.76
No. 43	0.49	0.51	0.32	No. 28	0.75	0.75	0.81
No. 65	0.50	0.55	0.10	No. 55	0.77	0.77	0.73
No. 46	0.52	0.52	0.51	No. 34	0.77	0.78	0.72
No. 15	0.52	0.53	0.46	No. 5	0.78	0.77	0.87
No. 19	0.55	0.56	0.53	No. 12	0.79	0.79	0.74
No. 22	0.55	0.55	0.56	No. 8	0.80	0.80	0.78
No. 25	0.58	0.57	0.75	No. 52	0.80	0.80	0.81
No. 18	0.61	0.60	0.67	No. 44	0.81	0.81	0.79
No. 32	0.62	0.62	0.57	No. 48	0.81	0.81	0.86
No. 4	0.62	0.63	0.57	No. 20	0.82	0.81	0.85
No. 66	0.62	0.63	0.59	No. 37	0.82	0.81	0.88
No. 35	0.63	0.63	0.70	No. 3	0.82	0.82	0.85
No. 26	0.64	0.63	0.65	No. 59	0.82	0.82	0.89
No. 21	0.64	0.65	0.60	No. 54	0.84	0.84	0.86
No. 13	0.66	0.67	0.62	No. 17	0.85	0.85	0.84
No. 1	0.66	0.65	0.75	No. 31	0.87	0.87	0.85
No. 11	0.67	0.68	0.58	No. 29	0.87	0.86	0.93
No. 6	0.67	0.67	0.68	No. 7	0.87	0.87	0.88
No. 27	0.67	0.67	0.69	No. 49	0.87	0.87	0.95
No. 39	0.68	0.68	0.66	No. 50	0.88	0.87	0.92
No. 10	0.68	0.68	0.62	No. 38	0.89	0.88	0.92
No. 64	0.68	0.68	0.68	No. 47	0.91	0.91	0.95
No. 30	0.69	0.69	0.67	No. 56	0.92	0.92	0.91
No. 62	0.70	0.69	0.80	No. 23	0.94	0.94	0.97
No. 36	0.70	0.71	0.66	No. 9	0.95	0.94	0.98
No. 58	0.71	0.70	0.80	No. 57	0.95	0.95	0.99
No. 42	0.71	0.71	0.75	No. 14	0.96	0.96	0.97
No. 16	0.72	0.72	0.66	No. 24	0.96	0.96	1.00