

2015

Assessing College Students' Quantitative and Scientific Reasoning: The James Madison University Story

John D. Hathcoat

James Madison University, hathcojd@jmu.edu

Donna L. Sundre

James Madison University, sundredl@jmu.edu

Mary M. Johnston

James Madison University, johnstmm@dukes.jmu.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/numeracy>

Recommended Citation

Hathcoat, John D.; Sundre, Donna L.; and Johnston, Mary M. (2015) "Assessing College Students' Quantitative and Scientific Reasoning: The James Madison University Story," *Numeracy*: Vol. 8 : Iss. 1 , Article 2.

DOI: <http://dx.doi.org/10.5038/1936-4660.8.1.2>

Available at: <http://scholarcommons.usf.edu/numeracy/vol8/iss1/art2>

Assessing College Students' Quantitative and Scientific Reasoning: The James Madison University Story

Abstract

Quantitative and scientific reasoning is a critical student learning outcome in higher education. Data are presented for large samples of undergraduate students who were assessed as entering freshmen and then again after completing 45-70 credit hours. Results are presented around four key issues that are central to educational assessment. First, entering freshmen with transfer credits for quantitative and scientific reasoning courses that fulfill general education requirements, on average, score similar to entering freshmen without such credit. About 97% of entering freshmen who had transfer credits received their credits through dual enrollment programs. As a sophomore-junior, students who had completed their general education requirements performed similar to students who had started, but not yet finished these requirements. Second, small to moderate correlations were observed between grade-point averages in relevant general education coursework and quantitative and scientific reasoning. Third, students' quantitative and scientific reasoning, on average, increases from freshmen to sophomore/junior years. Finally, the proportion of students who meet faculty-set standards substantially increases from pre-test to post-test. Taken together, results suggest that changes in quantitative and scientific reasoning are a function of relevant courses. Additional research is needed to examine the role of lower-level versus higher-level courses in student performance. Results also indicate a need to investigate how differences in the quality of dual enrollment courses facilitate quantitative and scientific reasoning.

Keywords

Quantitative Reasoning, Assessment, Accountability, Stakeholders, Scientific Reasoning

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

John D. Hathcoat is an Assistant Professor in Graduate Psychology and an Assistant Assessment Specialist in the Center for Assessment and Research Studies at James Madison University. His research interests include validity theory, instrument/scale development, performance assessment, and student outcomes assessment in higher education.

Donna Sundre is the Executive Director of James Madison University's (JMU) Center for Assessment and Research Studies (CARS) and has been at JMU for over 25 years. She is also a Professor of Graduate Psychology. Her research interests include quantitative and scientific reasoning, instrument development, validity, and examinee motivation in low-stakes testing conditions. Dr. Sundre is an associate editor of this journal.

Mary M. Johnston is an assessment consultant and doctoral candidate in the Assessment and Measurement Ph.D. program at James Madison University. She has served as a graduate assistant in the Center for Assessment and Research Studies. Her research interests include applying latent models to assess student

learning and development and examining the impact of motivation on testing performance across various contexts.

Introduction

Contemporary society is driven by data. Quantitative reasoning (QR), or what may be broadly defined as the use of higher-order reasoning to solve real-world numerical problems, is necessary to successfully navigate an ever-changing world. Consequently, numerous professional organizations, such as the Mathematical Association of America and the National Council of Teachers of Mathematics, have contributed to a growing movement calling for the establishment of QR as a student learning outcome. This movement is affirmed by the American Association of Colleges and Universities (2007) and more recently by the Lumina Degree Qualifications Profile (Lumina Foundation 2014), both of whom list QR as an intended consequence of higher education. Many other examples could be provided. Nevertheless, and despite the success of the QR movement in education, some reports indicate that many students fail to exhibit desired levels of QR proficiency (e.g., National Center for Education Statistics 2002; Adelman 2004). Such findings may in part arise from the fact that many students fail to engage in activities that promote QR development. For example, according to results that we pulled from the 2013 National Survey of Student Engagement report builder,¹ 20% of college seniors and 23% of incoming freshmen indicated that they ‘never’ used numerical information to examine a real-world problem or issue within an academic year. Such findings underscore the importance of monitoring QR development via assessment, as well as providing educational opportunities for students to advance these needed skills.

Scientific reasoning (SR) is an equally important student learning outcome. Advances in basic and applied science, such as technology and health care, suggest that all students minimally need an understanding of scientific inquiry. In fact, most institutions of higher education require science courses within their general education program (Warner and Koepfel 2009). At James Madison University² (JMU) the general education program aligns with our institutional mission *to prepare students to be educated and enlightened citizens who lead productive and meaningful lives*. Quantitative and scientific reasoning (QR-SR) is, in our view, one of the defining features of an educated and enlightened

¹ Custom reports from the 2013 National Survey of Student Engagement can be generated at the following website: http://nsse.iub.edu/html/report_builder.cfm. (This and other links in the text of this paper were accessed Nov. 29, 2014).

² JMU is a comprehensive public university located in central Virginia. JMU offers bachelor’s, master’s, and doctoral level programs and is classified by the Carnegie Classification system as a Master’s Colleges and Universities (larger programs). JMU annually enrolls approximately 18,400 undergraduates and 1,750 graduate students. Information about the current student body can be found at <http://www.jmu.edu/images/news/2014/2014-oct-factsheet.pdf>.

citizen. The Center for Assessment and Research Studies at JMU consists of both faculty and graduate students who are dedicated to using assessment results to fulfill this mission. We have studied QR-SR for over 20 years. Throughout this time we have refined our assessment methodologies (Sundre and Thelk 2010) and accumulated a growing body of evidence about student learning outcomes. This evidence also highlights issues with QR-SR that warrant further investigation in the context of higher education.

The purpose of the present article is to provide assessment results for QR-SR development among undergraduate students. Within this paper, four key questions are addressed about student learning outcomes that are central to educational assessment (see Pieper et al. 2008). These questions can be succinctly summarized: (1) Do students who participated in designated QR-SR courses have higher QR-SR scores than students who have not participated in such courses? (2) What is the relationship between QR-SR scores and course grades? (3) Does QR-SR reasoning change over time? and (4) Do students meet faculty expectations?

Prior to presenting our findings, the first section introduces the general education program at JMU and its assessment. This is followed by a brief description of the Natural-World 9 (NW-9), an assessment instrument composed of two subscales reflecting quantitative and scientific reasoning developed collaboratively by the faculty who teach our general education science and mathematics courses and faculty assessment liaisons in the Center for Assessment and Research Studies. Finally, we conclude by highlighting directions for subsequent research.

General Education Assessment at James Madison University

JMU's current General Education program was implemented in 1997 and is designed to help "students develop their ability to reason and make ethical choices; to appreciate beauty and understand the natural and social worlds they live in; to recognize the importance of the past and work towards a better future." The program is divided across five content clusters. Students' QR and SR skills are specifically addressed by Cluster 3, also known as the Natural World Cluster, which represents a mixture of fields including (but not limited to) physics, biology, chemistry, mathematics, psychology, and geography. JMU requires all students to complete at least 10-12 credit hours of coursework related to Cluster 3 prior to graduation.

Integral to JMU's General Education assessment model is the implementation of Assessment Day, a biannual university-wide event that is designed to assess students' knowledge, growth, and development over the course of their tenure at JMU. Specifically, students are required to participate in Assessment Day once as

entering freshmen prior to beginning classes in the fall (pre-test). This Assessment Day is fully integrated with a four-day required orientation for all new entering students. Students again participate in a spring Assessment Day when they have accumulated 45-70 credit hours; this represents the mid-point of a student's undergraduate academic career (post-test). They have achieved sophomore-junior status. Classes are canceled on the day of the spring administration to ensure student participation; therefore, there are no room or time conflicts. Students who do not attend their testing session are required to attend a make-up session. Further, registration for the next semester courses is blocked until a student fulfills the assessment requirements.

For all Assessment Day activities, students are randomly assigned to testing rooms with a variety of test configurations using the last two digits of their student ID. All students complete a battery of tests consisting of at least one cognitive (e.g., general education knowledge) and one non-cognitive (e.g., achievement motivation, academic self-efficacy) test administered by trained proctors. In other words, some students may be randomly assigned to take a history test whereas other students are randomly assigned to complete a quantitative and scientific reasoning test. Since student ID numbers do not change, it is possible to assure that students complete the same set of tests at both the entering-first-year and sophomore-junior test administrations. That is, to the extent possible, the set of tests students complete as freshmen is identical to the set they complete as sophomore-juniors. This pre-post design allows for true repeated measures and is one of the major strengths of the JMU's General Education assessment design.

The Natural World 9

Included in the Assessment Day tests is the Natural World-9 (NW-9), a 66-item test developed by JMU faculty and the Center for Assessment and Research Studies to assess Cluster 3's QR and SR learning objectives³ (Table 1; Sundre and Thelk 2010). Specifically, the NW-9 is composed of two tests: QR-9 and SR-9. The SR-9 consists of 49 items assessing six student learning objectives (1-4 and 7 and 8). The QR-9 consists of 26 items aligned with Cluster 3 student learning objectives 5 and 6 (Table 1), which are associated with quantitative skills and competencies. JMU faculty deemed quantitative reasoning to be a component of scientific reasoning; that is, they believe that students cannot engage in scientific reasoning without quantitative reasoning. Thus, the sum of all items is representative of scientific reasoning (QR+SR), as defined; this total is referred to

³ Cluster 3's learning objectives can be viewed at <http://www.jmu.edu/gened/about/Cluster%20Three%20Requirements.shtml>.

as the NW-9. The QR-9, SR-9 and NW-9⁴ have been extensively studied and are all currently in their ninth versions (see Sundre and Thelk 2010; Hurney et al. 2011). For clarity, we will refer to these tests as the QR and the QR+SR in this paper.

Table 1
Quantitative and Scientific Reasoning Student Learning Objectives

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

We return to the four questions outlined in the introduction regarding student learning. An affirmative response to any single one of them would not be sufficient evidence of quality educational impact; however, taken together, such evidence builds a case for educational quality. The following section will detail the answers to the four analytical questions we pose for every student cohort we have assessed.

Quantitative and Scientific Reasoning Assessment Results

Differences: Do students who participated in designated QR-SR courses have higher QR-SR scores than students who have not participated in such courses?

If students are meeting our stated learning objectives, their QR and QR+SR scores should improve as they complete Cluster 3's QR-SR-related coursework. Our assessment design allows us to review student exposure to relevant QR-SR coursework (i.e., Cluster 3 coursework) at two key time points: upon entry to the university as freshmen and at the mid-point of their undergraduate career. With respect to student differences, there are two questions that are of interest: (1) Do

⁴ The QR-9, SR-9, and NW-9 are commercially available at www.madisonassessment.com.

entering freshmen with relevant QR-SR course credit have on average higher QR and QR+SR scores than entering freshmen without such credit and (2) Do students with additional QR-SR coursework tend to score higher than students who have not participated in such coursework? As discussed below, although we have evidence that QR and QR+SR appears to be a function of relevant coursework, our evidence strongly suggests that it matters where and how this coursework was completed.

The first question in this section addresses whether entering freshmen with relevant QR-SR collegiate course credit have higher scores than students who are admitted without such course credit. Table 2 provides QR and QR+SR scores by type of relevant course credit for entering fall 2011 first-year students. These results indicate two interesting findings. First, students entering with relevant course credit achieved via Advanced Placement or International Baccalaureate (AP/IB) tests tend to perform higher than students without relevant course credit or students with relevant coursework achieved via transfer credit. Second, entering freshmen with relevant Cluster 3 transfer credits score similarly to students without course credit. This is a cause for concern that is reinforced by examining the zero-order correlations between the number of course credit hours received and students' performance (Table 3). The number of AP/IB credits has small zero-order correlations with QR scores, and, though the correlation between AP/IB credits and QR+SR scores was not statistically significant in the fall 2011 cohort, the magnitude of this correlation has been higher in previous years. Additionally, transfer credits are generally unrelated to student performance, and this finding has been consistent across numerous years of administration.

Table 2
Fall 2011 entering student QR and QR+SR average raw and percent correct scores by type of relevant QR-SR credit received outside of JMU

Credit type	N	QR		QR+SR	
		Raw Score, Mean (SD)	% Correct, Mean (SD)	Raw Score, Mean (SD)	% Correct, Mean (SD)
No credit	1400	17.25 (3.54)	66.33 (13.61)	45.79 (7.03)	69.37 (10.65)
AP/IB	156	20.33 (2.63)	78.21 (10.12)	52.25 (5.46)	79.17 (8.27)
Transfer	149	17.29 (3.30)	66.49 (12.71)	46.11 (6.51)	69.86 (9.87)

Note: QR = quantitative reasoning; SR = scientific reasoning; AP/IB = Advanced Placement/International Baccalaureate. Tests for significance indicate that students AP/IB course credit score higher than the other two groups. Students with transfer credit score similar to students without course credit.

Table 3
Fall 2011 correlations and r^2 for number of QR-SR credit hours completed and QR and QR+SR scores

Source of QR-SR Credit Hours	N	QR		QR+SR	
		r	r^2	r	r^2
AP Credit Hours	156	0.17*	0.031	0.14	0.019
Transfer Credit Hours	149	0.09	0.007	0.07	0.005

Note: In Fall 2010 the correlation between AP and QR+SR was .26 ($p < .01$).

* $p < .05$.

Accounting for these results across the different entering-student credit groups (i.e., no credit, AP/IB, and transfer) has been challenging, although there are a few possible explanations that arise when examining the students' characteristics (Table 4). First, the average age of students within each group is roughly 18.40 so such differences displayed above in the QR and QR+SR scores are not necessarily attributable to age. This result is not unexpected, however, because 97% of the 149 entering 2011 freshmen students with transfer credits received such credit through dual enrollment (i.e., college credit received by students while still enrolled in high school). However, differences in QR and QR+SR scores may be attributable to differences in entering students' ability/aptitude. That is, entering students with AP/IB credit had higher ability/aptitude scores as measured by the SAT Math, Verbal, and Writing, on average, than entering students with transfer or no credit (Table 4). In addition, entering students with transfer credit had similar SAT scores as entering students without such credit. Though differences in ability/aptitude may account for the fact that students with AP/IB credit tend to score higher on the QR and QR+SR than the other two groups, it is still a concern that entering students with relevant QR-SR transfer credits (i.e., Cluster 3 credits), on average, fail to score higher than entering students without course credit. This issue has been debated by our faculty across our science and mathematics departments, as well as in our admissions office. Our research, combined with faculty experiences, has called into question the value of some transfer and dual-enrollment credit-bearing experiences. Bear in mind these results relate only to entering first-year students.

Table 4
Fall 2011 demographics by type of relevant QR-SR credit outside of JMU

	No Credit (N=1400)	AP/IB Credit (N=156)	Transfer Credit (N=149)
<i>Gender, N (%)</i>			
Female	851 (60.8)	75 (48.0)	93 (62.4)
Male	548 (39.1)	81 (51.9)	56 (37.6)
<i>Age, Mean (SD)</i>			
	18.44 (0.41)	18.42 (0.34)	18.49 (0.40)
<i>Race, N (%)</i>			
American Indian	20 (1.4)	1 (0.6)	--
Asian	106 (7.6)	9 (5.8)	1 (0.7)
Black	69 (4.9)	1 (0.6)	3 (2.0)
Hispanic	42 (3.0)	5 (3.2)	--
Not Hispanic	478 (34.2)	54 (34.6)	71 (47.7)
Not Specified	35 (2.5)	5 (3.2)	5 (3.4)
Pacific Islander	3 (0.2)	--	--
White	646 (46.2)	81 (51.9)	67 (45.0)
<i>SAT Scores, Mean (SD)</i>			
Math	572.72 (63.60)	648.91 (66.29)	566.46 (59.38)
Verbal	569.84 (71.62)	620.41 (58.64)	560.28 (69.25)
Writing	564.16 (68.33)	603.54 (65.91)	553.54 (77.82)

Note: Students with AP/IB credit have higher SAT scores than both other groups. Students with transfer credits have similar SAT scores as entering students without course credit.

The second question addressed within this section pertains to whether QR and QR+SR scores increase as students take additional relevant QR-SR coursework during their JMU career. Presented in Table 5 are QR and QR+SR scores for the spring 2013 cohort of sophomores/juniors by the number of related QR-SR courses completed at JMU. Although very few students failed to complete any relevant coursework by their sophomore/junior year, these students had lower QR and QR+SR scores than students with relevant coursework (Table 5). These results are expected because the general education coursework should contribute to QR and QR+SR scores. However, as seen in Table 5, students appear to perform similarly irrespective of the number of relevant QR-SR credit hours they have completed. Although not reported here, this has been a consistent finding in recent years that has prompted additional analyses. Previous research has suggested that gains in QR may be a function of the type of course that is taken, in that gains may not be found in lower-level courses but are instead found as students take a sequence of lower- and upper-level courses (e.g., Grawe 2013). Put differently, it may be unreasonable to expect to see gains in QR and QR+SR across a series of courses that are designed to teach these skills at a basic level.

Table 5
Spring 2013 QR and QR+SR scores by number of Cluster 3 courses taken

# of Cluster 3 credits	N	QR		QR+SR	
		Raw Score, Mean (SD)	% Correct, Mean (SD)	Raw Score, Mean (SD)	% Correct, Mean (SD)
0	46	16.78 (4.21)	64.55 (16.17)	45.89 (8.10)	69.53 (12.28)
1	177	18.58 (3.40)	71.47 (13.08)	49.06 (7.17)	74.34 (10.87)
2	307	18.35 (3.58)	70.56 (13.79)	49.01 (7.33)	74.25 (11.11)
3 or more	477	18.74 (3.81)	72.06 (14.67)	49.54 (7.76)	75.06 (11.76)

Note: Tests for significance indicate that students with course credit have similar scores. However, students with credit tend to have higher scores than students with no relevant credits.

This insight aligns with results from a Science, Technology, Engineering, and Mathematics competition that we hold each year for graduating seniors in various disciplines. Academic programs volunteer to participate in the competition. For this competition, all graduating seniors from the Psychology, Biology, Chemistry, and Physics departments have been tested on QR and QR+SR. The department with the highest average score is deemed the winner of the competition and is awarded a plaque. Although we lack true repeated measures for these students because many of the participants were not randomly assigned to take the QR and QR+SR test as entering freshmen, or were not tested as sophomores/juniors because they did not have 45-70 credit hours, it is interesting to compare their QR and QR+SR scores to students who have completed their relevant coursework by their sophomore/junior years. For example, students with three or more relevant QR-SR course credits by their sophomore/junior year correctly answered 72.06% ($SD = 14.67$) of the QR items, whereas graduating Physics students correctly

answered 91.11% ($SD = 6.66$) of the QR items, on average. These results are also similar for graduating seniors from the other QR-SR-related disciplines. Though these results are not surprising, they are, at least minimally speaking, consistent with the position that cumulative growth in QR-SR may require courses taught beyond a basic level. Additional research is needed to investigate this possibility within the general education program.

Relationship with Course Grades: What is the relationship between QR-SR scores and course grades?

The second key question to address QR-SR general education assessment is answered by examining the relationship between QR and QR+SR scores and QR-SR course grades. That is, if our courses are instructing students on the QR-SR objectives, then their test scores should be positively associated with their Cluster 3 coursework grades. Although course grades reflect a combination of class activities that may not be necessarily related to student learning outcomes (e.g., attendance, timely submission of assignments), grades broadly indicate student achievement. Given that no single course is expected to do the work of the entire QR-SR area requirement, JMU students are required to complete a minimum of three courses and a lab (between 10-12 total credit hours) to fulfill our QR and SR requirements. However, examination of the correlations between QR-SR scores with individual QR-SR course grades should indicate that QR and QR+SR performance has a small to moderate relationship with QR-SR course performance. We would certainly be discouraged by any negative or zero correlations.

As expected, performance on the QR and QR+SR exhibits small to moderate positive correlations with Cluster 3 course grades. For example, when looking at the last three post-test administrations (Spring 2012-Spring 2014) of the QR and QR+SR, the correlations with Cluster 3 courses range from $r = .15$ to $r = .20$ for QR and QR+SR. Occasionally, we see higher correlations in the .30 to .45 range, but the important observation is that we do *not* see any negative or zero correlations, despite restriction of range for grades in a number of courses (such as a few teacher education courses). Thus, these correlations provide evidence that increased student achievement within QR-SR-relevant courses, as broadly defined by course grades, is associated with increases in QR and SR development.

Change over Time: Does QR-SR reasoning change over time?

The third central question to address QR-SR general education pertains to whether students are changing over time. If the QR-SR student learning objectives are achieved, students' QR and QR+SR scores should increase over time. Minimally, this requires a pre-test/post-test design that should explicitly focus on

students who have completed their QR-SR-relevant coursework requirements. Recall that our Assessment Day design provides us with true repeated measures for our student samples. Within this section, we present changes in QR-SR reasoning for two entering freshmen cohorts who finished their coursework requirements at post-test (i.e., in the spring of their sophomore or junior year). Specifically, Tables 6 and 7 present the pre-test and post-test results for the 2010-2012 and the 2011-2013 student cohorts, by course-completion status, for QR and QR+SR scores, respectively.

Table 6
Average QR scores by Cluster 3 course-completion status across time and cohort

Cluster 3 Course Status	2010 to 2012				2011 to 2013			
	<i>N</i>	Pre-test, Mean (<i>SD</i>)	Post-test, Mean (<i>SD</i>)	<i>d</i>	<i>N</i>	Pre-test, Mean (<i>SD</i>)	Post-test, Mean (<i>SD</i>)	<i>d</i>
Overall	761	17.66 (3.47)	18.83 (3.59)	0.32	867	17.69 (3.49)	18.90 (3.50)	0.48
Partial Completers	560	17.61 (3.43)	18.71 (3.51)	0.31	661	17.49 (3.55)	18.70 (3.57)	0.36
Completers	184	17.90 (3.61)	19.34 (3.65)	0.41	203	18.33 (3.23)	19.54 (3.22)	0.37

Note: Overall = all students who completed the pre-test and post-test; Partial Completers = students who had started but not yet completed the required Cluster 3 courses at post-test; Completers = students who had completed the Cluster 3 courses at post-test. All paired t-tests were statistically significant, $p < .001$.

Table 7
Average QR+SR scores by Cluster 3 course completion status across time and cohort

Cluster 3 Course Status	2010 to 2012				2011 to 2013			
	<i>N</i>	Pre-test, Mean (<i>SD</i>)	Post-test, Mean (<i>SD</i>)	<i>d</i>	<i>N</i>	Pre-test, Mean (<i>SD</i>)	Post-test, Mean (<i>SD</i>)	<i>d</i>
Overall	761	46.78 (6.57)	49.91 (7.17)	0.42	867	46.84 (6.72)	50.07 (7.17)	0.67
Partial Completers	560	46.74 (6.47)	49.79 (6.98)	0.42	661	46.39 (6.78)	49.70 (6.82)	0.55
Completers	184	47.04 (6.83)	50.53 (7.47)	0.46	203	48.31 (6.33)	51.28 (6.67)	0.52

Note: Overall = all students who completed the pre-test and post-test; Partial Completers = students who had started but not yet completed the required Cluster 3 courses at post-test; Completers = students who had completed the Cluster 3 courses at post-test. All paired t-tests were statistically significant, $p < .001$.

The results presented in Table 6 indicate students scored higher on the QR post-test when compared to their pre-test scores. We categorized the students into three groups: those who had started their Cluster 3 course requirements but had not completed them (partial completers); those who had completed their Cluster 3 courses (completers); and all students who completed the pre-test and post-test (overall). Note, we did not examine the performance of students who had not completed any Cluster 3 coursework by the post-test as the number of students in this group tends to be prohibitively small (i.e., less than five), thus disallowing accurate inferences. Paired t-tests were conducted, and all tests were significant ($p < .001$); more importantly, the effect sizes (Cohen's d) were all above .30 and ranged as high as .48 for all students entering in fall 2011. Cohen (1988) suggested benchmarks for interpreting effect sizes as follows: a "small" effect size is .20, a "medium" effect size is .50, and a "large" effect size is .80 or above. However, it should be noted that Cohen advised that these benchmarks may be

different for different fields of study. The assessment of general education outcomes is without question one of the more difficult areas of study, because the goals and objectives are by definition general. We contend that the effect sizes observed for our QR-SR outcomes are not small, and they easily qualify as medium effects. The results displayed in Table 6 are consistent with our previous findings and suggest that our students are showing both statistically significant and meaningful growth in QR over time.

Table 7 provides parallel results for the same cohorts on the QR+SR test score growth. Again, all paired t-tests resulted in statistically significant change over time ($p < .001$). Again, all paired t-tests resulted in statistically significant change over time ($p < .001$). The effect sizes presented were even larger than those observed for QR. The lowest effect size was .42 and the highest was .67. This indicates that the growth observed in QR+SR was two-thirds of a standard deviation increase. We believe these differences are educationally significant.

Meeting faculty standards: Do students meet faculty expectations?

The fourth key question to address QR-SR general education pertains to whether students are meeting faculty expectations. One of the strategies we have employed to foster Cluster 3 faculty engagement in general education assessment is to involve them in establishing a community expectation (i.e., standard) for student performance. In 2008, a modified Angoff (1971) standard-setting procedure was conducted with our Cluster 3 faculty who teach QR and SR ($n = 37$) to estimate the performance they would associate with a minimally competent student who had completed our QR and SR course requirements. The results of this process led to faculty deciding that a “minimally competent” student would answer 19 out of 26 QR items correct (i.e., 75%). A similar student is expected to correctly answer about 50 out of 66 QR+SR items correct (i.e., 76%). Incidentally, we believe our faculty established very high standards, as this test is a very arduous reasoning test.

If our QR-SR objectives are achieved, then the proportion of students who meet faculty standards should be higher at post-test than at pre-test. Further, a minimal number of students should be able to achieve this standard at entry to the university as freshmen. Table 8 provides information about the proportion of students meeting faculty standards by test administration and Cluster 3 coursework-completion status across five pre-post administrations of the QR and QR+SR. Specifically, coursework-completion status is decomposed into students who have started but not yet completed their Cluster 3 requirements at post-test (partial completers) and students who have completed Cluster 3 requirements at post-test (completers). The results in Table 8 are restricted to only those students who completed both the pre-test as entering freshmen and the post-test as

sophomores or juniors. As previously mentioned, we observed very few students who have not completed any QR-SR related courses at post-test. As a result, these students are not included in the present analysis.

Table 8
QR and QR+SR test performance compared with faculty expectations, by administration and Cluster 3 coursework completion status

Post-test Year	Pre-test Year	Test	Proportion Meeting Standard		
			Pre-test	Partial Completers	Completers
2009 (N = 1095)	2007	QR	.19	.37	.42
		QR+SR	.16	.42	.47
2010 (N = 1062)	2008	QR	.19	.38	.42
		QR+SR	.16	.40	.47
2011 (N = 1048)	2009	QR	.33	.43	.51
		QR+SR	.33	.48	.51
2012 (N = 761)	2010	QR	.30	.42	.51
		QR+SR	.30	.49	.54
2013 (N = 867)	2011	QR	.34	.47	.56
		QR+SR	.32	.50	.58

Note: QR = quantitative reasoning; QR+SR = scientific reasoning; Pre-test = students who met the standard as entering 1st-year freshmen at pre-test; Partial completers = students who started, but have not completed QR-SR Cluster 3 course requirements at post-test; Completers = students who finished the Cluster 3 course requirements at post-test.

Results in Table 8 display positive trends across five pre-test/post-test administrations of the QR and QR+SR. Across each year, the proportion of students who meet Cluster 3's faculty standards increased at post-test. For example, 34% of entering first-year students met faculty standards for QR on the pre-test administered in 2011. In contrast, 47% of partial coursework completers and 56% of coursework completers met faculty standards for QR at the post-test administered in 2013. The differences in the proportion of students who meet faculty expectations across time was statistically significant for both partial completers and completers ($z = 11.94, p < .001$; $z = 2.56, p < .001$, respectively). Similar patterns exist for the proportion of students meeting faculty expectations for QR+SR. For example, in 2013 about 50% of students who had partially completed their Cluster 3 requirements met faculty expectations for QR+SR at post-test, which is an increase from the 32% of students who met expectations at the pre-test ($z = 9.96, p < .001$). The proportion of students meeting faculty standards increased by 26% from the pre-test for Cluster 3 completers in 2013.

Another interesting finding observed in Table 8 is a rather dramatic increase in pre-test performance among entering freshmen over time. In 2008, only 16-19% of entering freshmen met faculty expectations at pre-test. However, in 2009, 33% of entering freshmen met faculty expectations for performance on QR and QR+SR. At the time, we suspected that this increase may be an aberration because our institution did not see any changes in selectivity or entering SAT scores. The pattern has been consistent, however, and so we now see it as a positive indication that *something important* has happened to our students prior to their arrival at our institution. It is not clear to us what would account for these findings. It may be that the Commonwealth of Virginia's late 1990's creation of the Standards of Learning,⁵ through which achievement expectations have been established for English, mathematics, science, and history/social science, is, indeed, having an impact on student learning. Nevertheless, despite increases in the proportion of students who are meeting faculty expectations as entering freshmen, it is evident from Table 8 that this proportion continues to increase at post-test. We continue to enhance student learning through our curriculum.

Summary and Conclusions

Our story of assessing QR-SR embodies many successes, future aspirations, as well as challenges that must be addressed if we wish to fulfill our mission to prepare educated and enlightened citizens within the general education program. As previously mentioned, QR-SR is a critical student learning outcome in higher education whose importance is reinforced by the demands of an increasingly complex society. Demands for accountability, coupled with skepticism among some about the value of higher education (e.g., Conner and Rabovsky 2011), acts to reinforce the importance of establishing effective assessment processes to investigate student learning outcomes. A strength of JMU's assessment design is that institutional structures are in place, such as Assessment-Day, that allow for monitoring of changes in student learning over time. Having such structures has also highlighted issues of concern that stand in need of subsequent research. We have suggested that an argument for educational quality, or impact, can be made by accumulating evidence centered on four questions. These four questions pertain to differences, relationships, change, as well as standards. Within this section we will briefly summarize these findings and underscore directions for future research.

Taken together, results from the four key questions imply that students tend to increase in QR-SR with relevant coursework. Evidence for this improvement is provided by positive correlations between student performance in relevant courses

⁵ Information about Commonwealth of Virginia's standards of learning may be found at the following website: <http://www.doe.virginia.gov/testing/index.shtml>

and QR-SR scores, increases in the proportion of students who meet faculty standards at post-test, as well as evidence indicating that QR-SR increases across time. Some caveats are necessary here. For example, although there are correlations between performance in relevant courses and QR-SR scores, in recent years we have failed to find expected cumulative increases in QR-SR scores with additional coursework. To put this differently, students who have completed their Cluster 3 coursework tend to perform similarly to students who have completed some, but not all, of such courses. It is possible that this finding is due to the fact that lower-level courses are taught at a basic level (Grawe 2013). Additional research is needed to investigate the conditions, if any, under which this finding may be moderated. For example, some courses are more aligned with specific objectives than are others. Cumulative increases in student performance may be anticipated for students taking courses that fail to substantially overlap in their coverage of Cluster 3 objectives. For example, if a specific course covered objectives 1 and 2, then cumulative growth would be anticipated if students later enrolled in a course focusing on objectives 3 and 4. A failure to find cumulative increases in student performance may be indicative of substantial overlap between objective coverage across Cluster 3 courses.

This issue of failing to see differences in QR-SR scores by either full or partial coursework completion is related to a second point of concern that is currently being discussed by faculty who teach general education courses. We have previously reported a substantial increase in the proportion of students who meet faculty standards at post-test. This is promising evidence of educational quality. Nevertheless, we believe that we can do much better. Faculty members have raised concerns about the proportion of students who are meeting standards at post-test, despite the increases. Recall that the faculty-set standards were estimated by specifying the performance of “minimally competent” students who have completed their relevant coursework. However, in 2013, 58% of students who completed their Cluster 3 requirements met these expectations. Many faculty members who teach these courses believe that this value should be higher. Discussions are currently underway about what this value should be, as well as what interventions may be necessary to increase the proportion of students who are meeting faculty expectations after they have completed their coursework. For example, we have recently sought to strengthen partnerships with pedagogical experts, faculty members, and assessment specialists to work in collaboration to increase QR-SR scores.

Finally, transfer students (97% of whom received credit via dual enrollment) score similarly on the pre-test as students who have no relevant coursework. This is an area of concern. If students receive credit for courses that meet stated objectives, then these students should, at least on average, have the knowledge and skills necessary to outperform students who have not yet received course

credit. Dual enrollment occurs when high school students earn college credit and high school credit simultaneously by enrolling in approved courses. In the state of Virginia, high school instructors must be certified to teach these courses, which in part requires that they have credentials that are similar to instructors at a community college. Dual enrollment may provide advantages to many students. For example, some research suggests that this may ease the transition to college (Bailey et al. 2003) and that students who participate in such programs tend to have higher college attainment than students who did not participate (Karp and Hughes 2008). Our findings do not contradict this research. However, our findings do raise concerns about variation in the quality and rigor of dual enrollment courses. Such concerns are supported by research indicating that the advantages of dual enrollment, at least for many students, may depend upon the type of course for which they received credit (Speroni 2011). Although our findings have ignited conversations on campus, as well as conversations with partner institutions, additional research is needed to investigate whether similar results are obtained at other institutions of higher education. Such research may have implications on public policy, such as requirements to certify instructors, curricular requirements for specific courses, or giving students credit for demonstrating skills as opposed to “seat-time” within a classroom.

In conclusion, we have presented promising evidence for QR-SR development among students at our institution. Despite the challenges that were raised, we remain optimistic about the future. Clearly, promoting QR-SR development requires collaboration from content and pedagogical experts who coordinate their efforts with assessment specialists with advanced training in methodology. Doing assessment well is consequently one piece of a much larger puzzle. Educational assessment provides opportunities to document and communicate successes while pinpointing substantive areas for continual improvement. Assessment, if done well, is a scholarly activity. As with other areas of scholarship, advancements in knowledge often lead to new, yet increasingly nuanced, questions that require further investigation. The results presented here, although evidence of particular successes, also point to areas that stand in need of additional attention. We are hopeful that other stakeholders who value QR-SR will join us in our efforts to advance both the assessment and facilitation of this important student learning outcome.

Acknowledgments

We would like to acknowledge the support of administration, Cluster 3 faculty, and the Assessment-Day team at James Madison University for making this article possible. We would also like to acknowledge the role of various graduate

assistants in the Center for Assessment and Research Studies who provided support with both collecting and analyzing assessment data.

References

- Association of American Colleges and Universities. 2007. College learning for the new global century: A report from the National Leadership Council for Liberal Education and America's Promise. http://www.aacu.org/leap/documents/GlobalCentury_final.pdf (accessed April 10, 2012).
- Adelman, C. 2004. Global preparedness of pre-9/11 college graduates: What the US longitudinal studies say. *Tertiary Education and Management* 10: 243–260. <http://dx.doi.org/10.1080/13583883.2004.9967130>
- Angoff, W. H. 1971. Scales, norms, and equivalent scores. In *Educational measurement*, ed. R. L. Thorndike, 508–597. Washington, DC: American Council on Education.
- Bailey, T. R., K. L. Hughes, and M. M. Karp. 2003. *Dual enrollment programs: Easing transitions from high school to college*. CCRC Brief. <http://files.eric.ed.gov/fulltext/ED475805.pdf> (accessed September 1, 2014).
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- Conner, T. W. and T. M. Rabovsky. 2011. Accountability, affordability, access: A review of the recent trends in higher education policy research. *Policy Studies Journal* 39: 93–112. http://dx.doi.org/10.1111/j.1541-0072.2010.00389_7.x
- Grawe, N. D. 2013. Does completion of quantitative courses predict better quantitative reasoning-in-writing proficiency? *Numeracy* 6 (2): Article 11. <http://dx.doi.org/10.5038/1936-4660.6.2.11> (accessed August 26, 2014).
- Hurney, C. A., J. Brown, H. P. Griscom, E. Kancler, C. J. Wigtil, and D. L. Sundre. 2011. Closing the loop: Involving faculty in the assessment of scientific and quantitative reasoning skills of Biology majors. *Journal of College Science Teaching* 40 (6): 18–23.
- Karp, M. M. and K. L. Hughes. 2008. Study: Dual Enrollment Can Benefit a Broad Range of Students. *Techniques: Connecting Education and Careers* 83: 14–17.
- Lumina Foundation. 2014. *The Degree Qualifications Profile 2.0*. Indianapolis, IN: Author.
- National Center for Education Statistics. 2002. *Adult literacy in America: A first look at findings of the national adult literacy survey*. 3rd ed. Washington, DC: U.S. Department of Education. <https://nces.ed.gov/pubs93/93275.pdf> (accessed August, 26 2014).

- National Survey of Student Engagement. 2013. NSSE 2013 U.S. Classroom Experiences by Gender. http://bl-educ-cprtest.ads.iu.edu/SAS/rb_nsse.html (accessed August 23, 2014).
- Pieper, S. L., K. H. Fulcher, D. L. Sundre, and T. D. Erwin. 2008. What do I do with the data now?: Analyzing assessment information for accountability and improvement. *Research and Practice in Higher Education* 2 (1): 4–10.
- Speroni, C. 2011. High school dual enrollment programs: Are we fast-tracking students too fast? An NCPR Working Paper. *National Center for Postsecondary Research*. <http://files.eric.ed.gov/fulltext/ED527527.pdf> (accessed February 10, 2014).
- Sundre, D. L. and A. D. Thek. 2010. Advancing assessment of quantitative and scientific reasoning. *Numeracy* 3 (2), Article 2. <http://dx.doi.org/10.5038/1936-4660.3.2.2> (accessed February 10, 2014).
- Warner, D. B. and K. Koeppel. 2009. General education requirements: A comparative analysis. *The Journal of General Education* 58: 241–248. <http://dx.doi.org/10.1353/jge.0.0050>