Numeracy
Advancing Education in Quantitative Literacy

Volume 6 | Issue 2 Article 11

2013

Does Completion of Quantitative Courses Predict Better Quantitative Reasoning-in-Writing Proficiency?

Nathan D. Grawe
Carleton College, ngrawe@carleton.edu

Recommended Citation

Grawe, Nathan D.. "Does Completion of Quantitative Courses Predict Better Quantitative Reasoning-in-Writing Proficiency?." *Numeracy* 6, Iss. 2 (2013): Article 11. DOI: http://dx.doi.org/10.5038/1936-4660.6.2.11

Does Completion of Quantitative Courses Predict Better Quantitative Reasoning-in-Writing Proficiency?

Abstract

Using data from Carleton College, this study explores the connection between students' completion of a range of quantitative courses and the quality of their quantitative reasoning in writing (QRW) as exhibited in courses throughout the undergraduate curriculum during the first two years of college. Because the assessment takes place in the context of a campus-wide initiative which has improved QRW on the whole, the study identifies course-taking patterns which predict stronger than average improvement. Results suggest QRW is not exceptionally improved by taking courses in statistics, principles of economics, or in the social sciences more broadly. QRW performance is, on the other hand, correlated strongly with having taken a first-year seminar specifically designed to teach QR thinking and communication. To a lesser degree, QRW is correlated with courses in the natural sciences and upper-level calculus. It is impossible to rule out all forms of selection bias explanations for these patterns. However, the broad pattern of correlations between QRW, courses, and standardized test scores argues for a causal interpretation.

Keywords

quantitative reasoning, writing, assessment

Cover Page Footnote

Nathan D. Grawe is Associate Professor of Economics at Carleton College. His involvement in quantitative reasoning has focused on assessment and faculty development.

Introduction

In the dozen years since *Achieving Quantitative Literacy: An Urgent Challenge for Higher education* (Steen 2004), the quantitative reasoning (QR) movement indeed has done much to address the gaps in teaching tools, community, and assessment (Grawe 2012). Specifically, teacher-scholars have developed new teaching activities, modules, and courses to provide students greater QR experience.[1] Through regular meetings and a journal, the National Numeracy Network has expanded its footprint as a disciplinary society addressing Steen's concern that no discipline "consistently call[s] to make quantitative literacy a priority of education at the college level" (Steen 2004, p. 15). And, most recently, assessment tools have been developed to measure student performance so that it can no longer be said that "[quantitative literacy] is largely absent from our current systems of assessment and accountability" (Steen 2004 p. 11). (See, for example, the tools discussed in Grawe, 2012, or the newly developed Quantitative Literacy and Reasoning Assessment.[2])

In this paper, I aim to build on this foundation by contributing to the limited, but growing literature that assesses the effectiveness of various approaches to QR instruction. James Madison University (JMU) (2012) has studied the connection between QR performance and natural science courses among more than 4000 students over the last five years. They evaluated student performance using the 23-item, multiple-choice "QR Test," which is designed to assess two learning outcomes:

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

They find a positive correlation between student QR Test scores and both the number of "Natural World" course credits earned and grades received in those courses. This finding suggests that student QR performance can be improved by persistent, intentional instruction.[3]

[1] A large collection of these activities, including those supported by the National Numeracy Network.

[2] Information about the QLRA can be found at http://serc.carleton.edu/qlra/index.html.

[3] While the QR Test was developed at JMU, it has been taken by thousands of students at institutions across the United States and abroad. More information about the test and how it can be purchased for assessment at other institutions can be found at http://www.madisonassessment.com/assessment-testing/quantitative-reasoning-test/ .

Boersma and Klyve (2013) note that explicit prompts like those found in multiple-choice contexts are limited in that they cannot evaluate whether students will seek out quantitative information when faced with open-ended problems. Those authors use an adaptation of the rubric developed in Boersma et al. (2011), itself an adaptation of the Valid Assessment of Learning in Undergraduate Education (VALUE) rubric created by the Association of American Colleges and Universities. They compare the performance of honors-program students with non-honors peers in a "Quantitative Reasoning in the Contemporary World" course. All students were asked to evaluate the arguments in a newspaper story at the beginning of the term. At the end of the term, students were given the same prompt in regard to a second newspaper article. The authors report few gains in either group from the beginning to the end of the course. They note, however, that their experimental design makes it impossible to rule out the possibility that the results are confounded by differences in the quantitative complexity (from a student perspective) in the two articles; students may be stronger at the end of the term but receive the same score due to the relative difficulty of the latter news story.

The present study contributes to this literature through transcript analysis of students at Carleton College. The goal is to uncover which course-taking patterns are correlated with stronger performance in quantitative reasoning in writing (QRW). As is detailed in the data and methods section which follows, this assessment is based on a reading of papers written throughout the curriculum during the first two years of college. Because students wrote their papers without any knowledge that they would be assessed for QRW, this instrument, unlike other assessment instruments, does not prompt students to engage in QRW practices. (Of course, some of the assignments to which the papers respond do explicitly call for QRW.)

Since 2004, Carleton's faculty has been engaged in a broad curricular initiative to enhance student QR across the curriculum. As Grawe (2011) details, the quality of QRW on the campus as a whole has improved. So, in that context, the question explored in this paper is closer to "Which course-taking patterns predict *the greatest* impact on QRW?" than "Which course-taking patterns predict an impact on QRW?"

The results section below shows that completing two or three terms of calculus predicts better QRW performance. By contrast, successful completion of "applied" courses such as introduction to statistics, principles of economics, or social sciences courses more broadly predict scores that are no higher. Completion of natural science courses predicts modest QRW gains. And, completion of a first-year seminar (FYS) designed specifically to teach quantitative thinking and communication predicts a four-fold increase in the rate of exemplary QRW performance.

The conclusion of the results section considers the possibility that the positive results can be explained by selection bias. The most obvious selection bias story is that "strong mathematics" students opt into natural science courses and the QR FYS. However, the fact that no correlation is found between QRW scores and courses in calculus, introduction to statistics, principles of economics, and social science courses argues against this interpretation. In addition, there is no correlation between QRW performance and SAT or ACT mathematics scores. Ultimately, it seems the data could support a finding similar to that reported in James Madison University (2012): intentional teaching of QR methods and practices can lead to improved student performance.

The paper concludes with a discussion of what these results can (and cannot) say about the utility of various courses in improving QR.

Data and Methods

The Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) assessment instrument used in this study is the same as that described in detail in Grawe et al. (2010). To review briefly, at the end of their sophomore year, all Carleton students are required to submit 3–5 papers previously written for one of their courses to demonstrate capacities designated by the college's writing program. The writing program evaluates these portfolios as "exemplary," "pass," or "needs work."

The eligible population for the present analysis is all students who received "exemplary" or "pass" scores from the writing program who also gave permission for their work to be used for research purposes (roughly 85% of the entire student population). From this population, in 2008 we assessed a random sample of 72 students in the graduating class of 2010. In 2009, we added to the dataset, evaluating portfolios of a randomly selected half of eligible students in the graduating classes of 2006 and 2011—approximately 200 students from each class. However, when studying the QR-specific first-year seminar, the sample sizes were so small that we oversampled students from this course and evaluated all available students in this subgroup. In 2010, we expanded our database further by assessing portfolios from all eligible students in the graduating classes of 2007 and 2012. For each student in our sample, we randomly selected one paper from the writing portfolio to be assessed for QRW. On average, the assessed papers were written 2.92 years before graduation—that is, at the very tail end of the first year.

Using the QuIRK rubric, faculty and staff readers coded these papers for relevance, extent, and quality of QRW. The relevance and quality items on the rubric are used in the analysis here. The former was coded into three possible categories (Table 1). QRW was deemed "centrally relevant" if the use of quantitative evidence "address[ed] a central question, issue, or theme" in the paper. An example of such a work might be a lab report that examines the frequency with which a genetic marker appears in two sub-populations of insects. QR was coded "peripherally relevant" if quantitative information would "provide useful detail, enrich descriptions, present background, or establish frames of reference." For example, if a student wrote a paper arguing for a particular philosophical definition of "poverty," she may choose to introduce the paper with reference to rising poverty rates in the United States. While there are many other, non-quantitative ways to introduce such an argument—quantitative evidence need not be central to this question of philosophical definition—the choice to place an empirical frame around the discussion sets QR in a peripheral and supporting role in this paper. Of course, a third possibility is that QR has no relevance to the paper.

Table 1
Sample Descriptive Statistics

	All Papers	By papers' QR relevance		
		Centrally Relevant	Peripherally Relevant	Not QR Relevant
	n = 1,038	*n* = 266	*n* = 223	*n* = 549
Academic Division				
Arts and Literature	33.1%	13.5%	28.8%	44.3%
Humanities	22.4%	9.0%	22.1%	29.1%
Natural Sciences	14.7%	40.6%	8.6%	4.6%
Social Sciences	24.4%	31.2%	32.9%	17.7%
Interdisciplinary	5.4%	5.6%	7.7%	4.4%
Rated QRW Quality				
Rating = 1 (low)	35.5%	18.9%	55.4%	-
Rating = 2	28.1%	31.3%	24.3%	-
Rating = 3	29.8%	38.1%	19.8%	-
Rating = 4 (high)	6.6%	11.7%	0.5%	-

QRW quality was coded on a four-point scale using rubric language described in detail in Grawe et al. (2010). Whether speaking of central or peripheral QR, the scores can be summarized as follows:

1. Student's use of QR (or failure to use QR) leads to significant problems in the paper.
2. Student's use of QR in support of the paper's goals is only partially effective.
3. Student's use of QR in support of the paper's goals is good.
4. Student's use of QR in support of the paper's goals is exemplary.

Two statistical methods are used to explore the correlation between course completion and QRW scores. The first is a Pearson chi-squared test which tests the observed data against a null hypothesis that QRW proficiency and course completion are independent. The associated *p*-value provides the probability of seeing the observed pattern in the sample given that, in the population, the two variables are independent. This is a weak test in this context because it is possible to imagine changes in the proficiency distribution which are not necessarily improvements. For example, suppose a course leads weak students to flounder even more while it prepares strong students to excel. The Pearson chi-squared test will uncover this difference, but it is not clear that anyone would describe the net change as improvement.

The Wilcoxon rank-sum test (sometimes called the Mann-Whitney U test) provides one way to measure the improvement we are interested in. It estimates, based on the observed distributions, the probability that a randomly selected student in one group will have a higher QRW score than a randomly selected student in another group. An associated *p*-value can be calculated to assess the hypothesis that this probability is 50% (i.e., testing the null hypothesis that there is no reason to think that students in one group will generally outperform students in the other).

Patterns of Course Taking and QRW Performance

Calculus and Statistics

Calculus is the most common means by which students fulfill Carleton's "Formal or Statistical Reasoning" requirement.[4] In our sample, 47.8% of the students took at least one of the three offered levels of calculus before writing the paper we assessed.[5] (Calculus III is Introduction to Multivariable Calculus.) Table 2 presents information on QRW proficiency with students divided based upon their calculus course-taking experience. Because the uses of central and peripheral QR are distinct, they are considered separately. The top panel of the table reports QRW proficiency in papers of central QR relevance while the lower panel reports proficiency in papers of peripheral relevance. The columns are ordered by the level of the course. Here and in subsequent tables, the statistical test reported at the bottom of each column relates to the comparison between the group in question and all those with lesser course experience. For example, in the case of Calculus II, those who have taken Calculus II (but not Calculus III) are compared with those who have either taken no calculus or completed only Calculus I.

[4] The students in our study were/are governed by a slightly different requirement. However, the prevalence of calculus to fulfill Carleton's "mathematics requirement" remains more or less unchanged.

[5] In addition, some students have completed AP calculus courses. Throughout this paper, completion of a course means completion of a course at Carleton.

Table 2
Distribution of QRW Proficiency by Calculus Completion

	No Calculus	Calculus I, but not Calculus II	Calculus II, but not Calculus III	Calculus III
Central Relevance				
Quality Score	$n = 147$	$n = 39$	$n = 32$	$n = 83$
1	24.5%	12.8%	9.4%	20.5%
2	31.3%	41.0%	31.3%	26.5%
3	35.4%	43.6%	37.5%	36.4%
4	8.8%	2.6%	21.9%	16.9%
Pearson chi-squared test *p*-value		0.18	0.05	0.32
Wilcoxon rank-sum probability (*p*-value)		0.52 (0.61)	0.62 (0.02)	0.54 (0.29)
Peripheral Relevance				
Quality Score	$n = 132$	$n = 32$	$n = 26$	$n = 56$
1	50.0%	59.4%	73.1%	55.4%
2	27.3%	12.5%	19.2%	21.4%
3	22.0%	28.1%	7.7%	23.2%
4	0.8%	0.0%	0.0%	0.0%
Pearson chi-squared test *p*-value		0.33	0.19	0.92
Wilcoxon rank-sum probability (*p*-value)		0.48 (0.65)	0.38 (0.03)	0.50 (0.99)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of Calculus II those who have taken Calculus II (but not III) are compared with those who have either taken no calculus or only completed Calculus I.

First, consider cases of central use of QR. The data provide little evidence that taking Calculus I is associated with better QRW among our students. However, those who have taken Calculus II and III appear more likely to receive “exemplary” scores of 4. (When students who have completed Calculus II and III are combined and compared with all others, the Pearson chi-squared test *p*-value is 0.03). By contrast, there is no evidence of improved peripheral QR as students go deeper into the Calculus sequence. The Calculus II students actually performed notably worse than those with less Calculus experience.

Many have argued that mathematics and QR can be differentiated in part by the former’s attention to abstraction and deductive argument in contrast to the latter’s emphasis on context and inductive reasoning. (Steen, 2004, represents one good example of this distinction.) As a result, statistics may be more closely related to QR than is calculus. Carleton offers three entry points for statistics. The first two (Statistics 115 and Statistics 215) are aimed at those who have not had multivariable calculus. These two courses differ in the level of treatment, but students who complete the former are not permitted to earn credit in the latter because of the large overlap in content.

The third option is an advanced two-course sequence: Probability (Statistics 265) paired with Introduction to Statistical Inference. In this paper, all those who complete the Probability course will be coded as having completed an introduction to statistics. Because only 12 students in the sample completed Probability before writing the paper that we assessed, these students will be grouped with those who completed Statistics 215.

The statisticians on our campus have been consistently involved with our broader quantitative reasoning initiative, and they are supportive of the way we have situated QR in the context of writing. This support is evident in the fact that most statistics courses include writing assignments. Of course, these courses also have a heavy content load. Because Statistics 115 and 215 serve as pre-requisites to methods courses in other departments, the statisticians cannot devote an inordinate amount of time to QRW at the expense of this content.

Table 3
Distribution of QRW Proficiency by Introduction to Statistics Completion

	No Intro Statistics	Statistics 115, but not 215 or 265	Statistics 215 or 265
Central Relevance			
Quality Score	n = 195	n = 83	n = 23
1	19.0%	19.3%	34.8%
2	32.3%	33.7%	13.0%
3	36.4%	38.6%	34.8%
4	12.3%	8.4%	17.4%
Pearson chi-squared test p-value		0.83	0.11
Wilcoxon rank-sum probability (p-value)		0.48 (0.65)	0.49 (0.84)
Peripheral Relevance			
Quality Score	n = 164	n = 73	n = 9
1	58.5%	46.6	55.6%
2	17.7%	34.3%	33.3%
3	23.2%	19.2%	11.1%
4	0.6%	0.0%	0.0
Pearson chi-squared test p-value		0.04	0.82
Wilcoxon rank-sum probability (p-value)		0.53 (0.33)	0.47 (0.75)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of Statistics 215 those who have taken Statistics 215 (but not 265) are compared with those who have either taken no introductory statistics or only completed Statistics 115.

In our sample, 33.3% of students completed one of these three entry points to statistics in terms up to and including the one in which they wrote the paper in our sample. Because these courses are clearly ordered in mathematical depth, Table 3 presents comparisons similar to those in Table 2—those with no statistics introduction (column 1), those who completed Statistics115 (column 2), and those who completed Statistics 215 or 265 (column 3). The statistical tests in columns 2 and 3 compare students in the column in question with all of those with less experience (i.e., in columns to the left). The results show no evidence that completion of introductory statistics is correlated with stronger QWR performance.

Quantitative Courses Outside the Mathematics Department

Given that QR can be understood as the application of mathematics skills in context, many institutions have adopted graduation requirements that give credit for, and sometimes even require, QR experience outside mathematics (examples in Hillyard 2012, p. 12-13). Here, I examine the degree to which completion of non-mathematics courses commonly seen as contributing to QR development is correlated with QRW proficiency. First, I examine principles of economics, which is perceived as one of the most quantitative introductory social science courses. For example, at Macalester College, which gives courses various "points" toward a quantitative reasoning requirement, principles of economics is one of only a few offerings that fulfill the requirement in a single course (Bressoud 2009). Next, I look at social science courses more broadly. Finally, I look at courses in the natural sciences as the James Madison University study found positive correlations between QR and completion of courses in their Natural World sequence.

At Carleton, principles of economics is divided into two courses, Principles of Microeconomics and Principles of Macroeconomics, which can be taken in any order. In our sample, 20.7% of students took one of these courses before or in the same term as the course for which they wrote the assessed paper, and an additional 9.0% took both. Table 4 summarizes the relationship between completion of these courses and QRW scores. The results consistently show no discernible difference between those who take economics and those who do not.

While principles of economics is often held out as particularly quantitative, all of the social sciences offer opportunities to use and explore quantitative evidence in the context of arguments. Nearly all students (75.3%) had taken at least one social science class in the terms up to and including the one in which they wrote the assessed paper.[6] Just over one-quarter took only one course; 19.9% completed two; and around 15% took four or more. The average number of social science courses completed was 1.85. Table 5 shows the relationship between QRW proficiency scores and social science course-taking in our sample. The pattern generally follows that seen in principles of economics. More course-taking in the social sciences does not predict better QRW assessment in either the peripheral or central QR context. There is even some weak evidence that taking many social science courses predicts more poor-quality papers of central QR relevance and fewer high-quality ones.

---

[6] Here the "social sciences" include economics, psychology, sociology, anthropology, political science, linguistics, and education studies.

Table 4
Distribution of QRW Proficiency by Principles of Economics Completion

	No Principles of Economics	One Course, but Not Both	Both Principles Courses
Central Relevance			
Quality Score	*n* = 197	*n* = 61	*n* = 43
1	20.8%	18.0%	20.9%
2	28.4%	39.3%	32.6%
3	27.6%	34.4%	37.2%
4	13.2%	8.2%	9.3%
Pearson chi-squared test *p*-value		0.38	0.97
Wilcoxon rank-sum probability		0.47	0.48
(*p*-value)		(0.42)	(0.72)
Peripheral Relevance			
Quality Score	*n* = 175	*n* = 51	*n* = 20
1	53.1%	64.7%	45.0%
2	23.4%	17.7%	35.0%
3	22.9%	17.7%	20.0%
4	0.6%	0.0%	0.0%
Pearson chi-squared test *p*-value		0.51	0.61
Wilcoxon rank-sum probability		0.44	0.44
(*p*-value)		(0.16)	(0.55)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of the middle column those who have taken either Principles of Microeconomics or Principles of Macroeconomics (but not both) are compared with those who have taken neither.

Table 5
Distribution of QRW Proficiency by Social Science Course Completion

	No Courses	One Course	Two or Three Courses	Four or More Courses
Central Relevance				
Quality Score	*n* = 65	*n* = 62	*n* = 119	*n* = 55
1	13.9%	24.2%	19.3%	25.5%
2	30.8%	22.6%	34.5%	34.6%
3	36.9%	43.6%	36.1%	30.9%
4	18.5%	9.7%	10.1%	9.1%
Pearson chi-squared test *p*-value		0.20	0.51	0.54
Wilcoxon rank-sum probability (*p*-		0.44	0.46	0.44
value)		(0.26)	(0.26)	(0.15)
Peripheral Relevance				
Quality Score	*n* = 49	*n* = 79	*n* = 72	*n* = 46
1	53.1%	51.9%	59.7%	54.4%
2	26.5%	25.3%	20.8%	19.6%
3	18.4%	22.8%	19.4%	26.1%
4	2.0%	0.0%	0.0%	0.0%
Pearson chi-squared test *p*-value		0.59	0.67	0.78
Wilcoxon rank-sum probability (*p*-		0.51	0.46	0.51
value)		(0.87)	(0.36)	(0.75)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of the third column those who have taken between two and four courses are compared with those who have taken either one or none.

Natural sciences courses are also commonly included among lists of QR application courses.[7] Table 6 examines the correlation between QRW proficiency and the number of courses completed in natural science departments up to the term in which the assessed paper was written. Just under 40% of students in our sample did not complete a natural science course on or before the term in which they wrote the paper selected for assessment. Approximately one-quarter of students completed only one natural science course and 12.0% completed four or more.[8] The mean number of courses completed is 1.44. Not surprisingly, students who declare a major in the natural sciences (25.1% of our sample) complete roughly double this number.

Table 6
Distribution of QRW Proficiency by Natural Science Course Completion

	No Courses	One Course	Two or Three Courses	Four or More Courses
Central Relevance				
Quality Score	n = 80	n = 99	n = 72	n = 50
1	28.8%	15.2%	15.3%	24.0%
2	36.3%	33.3%	34.7%	14.0%
3	22.5%	40.4%	41.7%	46.0%
4	12.5%	11.1%	8.3%	16.0%
Pearson chi-squared test p-value		0.04	0.43	0.04
Wilcoxon rank-sum probability		0.59	0.53	0.56
(p-value)		(0.03)	(0.45)	(0.19)
Peripheral Relevance				
Quality Score	n = 94	n = 64	n = 58	n = 30
1	57.5%	57.8%	53.5%	43.3%
2	27.7%	20.3%	19.0%	23.3%
3	13.8%	21.9%	27.6%	33.3%
4	1.1%	0.0%	0.0%	0.0%
Pearson chi-squared test p-value		0.40	0.33	0.37
Wilcoxon rank-sum probability		0.51	0.54	0.58
(p-value)		(0.77)	(0.35)	(0.12)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of the third column those who have taken between two and four courses are compared with those who have taken either one or none.

Table 6 presents the distribution of assessed QRW quality by number of natural science courses completed at the time the paper was written. When

[7] Here the "natural sciences" include biology, chemistry, geology, physics, and astronomy.

[8] These course counts include only classroom courses; they do not count lab sections associated with courses. It does include some courses of less-than-term length, however. So, if a student took two half-term courses, this would be counted as two courses rather than one. (The level of detail in the data made available to me did not permit distinguishing courses by number of credits.) Such courses are typically upper-level courses and are uncommon in our sample of courses taken in the first two years.

examining centrally relevant QR papers, we see that students who have completed more natural science courses are more likely to be given "good" or "exemplary" ratings and less likely to be given ratings in the bottom two quality categories. The effect sizes are sizable with the likelihood of a paper falling in the top two categories doubling as we move from no natural science courses to four or more. The data from peripherally relevant QR papers show a similar pattern, though of slightly smaller magnitude. The smaller sample sizes among peripheral QR papers make these differences statistically insignificant.

Specifically Designed QR Courses

With support from the National Science Foundation and the W. M. Keck Foundation, Carleton has revised or created courses with the intention of improving student QRW. The course revisions are spread throughout the curriculum and typically involve adding new, QR-focused assignments to courses already in the catalog.[9] Perhaps the most aggressive curricular revision was Neil Lutsky's new course titled "Measured Thinking," a first-year seminar which focuses entirely on QR.[10] Some of the revised courses were offered in multiple sections by multiple professors. A professor assigned to one section of these courses may have added QR while other professors teaching other sections of the same course number did not. In such multi-section cases, the course was coded as un-revised unless all of the professors teaching the course revised their section. In addition, many of these courses were offered to students in our sample both before and after revision. Only students who took the course after the revision are coded as taking a QR-revised course. The sample includes 18 different QR-revised courses in seven different departments. Just over 4% of students completed the Measured Thinking seminar and another 10.8% completed at least one other QR-revised course in a term up to and including the one in which they wrote the assessed paper.[11]

Table 7 examines the quality of QRW exhibited by students who had and had not completed one of the revised courses. The first column presents scores for students who took no QR-revised courses while the second column presents the distribution of scores for those who took a QR-revised course other than the Measured Thinking first-year seminar. The final column presents the score distribution for students who completed the Measured Thinking seminar. As in

---

[9] Descriptions of revised assignments and courses can be found at http://serc.carleton.edu/quirk/Infusing/index.html.

[10] A detailed description of the course and its syllabus and assignments can be found at http://serc.carleton.edu/quirk/courses/measured_thinking.html.

[11] Of the 216 students who took a QR-revised course but not the QR first-year seminar, 203 took one revised course and 16 took two revised courses. None took more than two revised courses.

the previous analyses, the Pearson chi-squared and Wilcoxon rank-sum tests compare each group with all groups presented to the left of it in the table. Because the sample sizes are small for QR first-year seminar participants, the table also presents results from combining the results of central and peripheral QR use.

Table 7
Distribution of QRW Proficiency by QR-Revised Course Completion

	No QR-Revised Courses	QR-Revised Course, but not QR First-Year Seminar	QR First-Year Seminar
Central Relevance			
Quality Score	*n* = 252	*n* = 35	*n* = 14
1	21.0%	17.1%	14.3%
2	32.1%	25.7%	28.6%
3	36.9%	45.7%	14.3%
4	9.9%	11.4%	42.9%
Pearson chi-squared test *p*-value		0.72	0.00
Wilcoxon rank-sum probability (*p*-value)		0.55 (0.32)	0.63 (0.09)
Peripheral Relevance			
Quality Score	*n* = 216	*n* = 25	*n* = 5
1	57.8%	36.0%	20.0%
2	21.3%	40.0%	20.0%
3	20.4%	24.0%	60.0%
4	0.5%	0.0%	0.0%
Pearson chi-squared test *p*-value		0.14	0.20
Wilcoxon rank-sum probability (*p*-value)		0.59 (0.09)	0.72 (0.06)
Central and Peripheral Relevance Combined			
Quality Score	*n* = 468	*n* = 60	*n* = 19
1	38.0%	25.0%	15.8%
2	27.1%	31.7%	26.3%
3	29.3%	36.7%	26.3%
4	5.6%	6.7%	31.6%
Pearson chi-squared test *p*-value		0.27	0.00
Wilcoxon rank-sum probability (*p*-value)		0.57 (0.07)	0.68 (0.01)

Note: For both the Pearson chi-squared and Wilcoxon rank-sum test, the group in question is compared with all those with lesser course experience. For example, in the case of exposure to a QR-revised course those who have taken a QR-revised course (but not the QR first-year seminar) are compared with those who have taken no QR course.

While completion of a QR-revised course does not substantially alter the quality pattern among centrally relevant papers, among peripherally relevant papers, students are about 40% less likely to fall in the lowest-quality category. The most impressive differences are found when comparing students who completed the QR first-year seminar with all others. Students from that seminar are four times as likely to be given the highest-quality rating for their use of quantitative evidence when writing papers for which QR is centrally relevant—a

33-percentage-point improvement. In papers of peripheral QR relevance, students who took the seminar are roughly three times as likely to receive a "good" score of 3, a 40-percentage-point improvement in the assessment of student quantitative writing. And the probability that a randomly selected seminar student will out-perform a randomly selected non-seminar student is estimated between two-thirds and three-quarters.

Several Potential Biases

The results above should be placed in the context of several potential biases. First is Carleton's institutional context. Carleton is a small, residential liberal arts college with about 525 students per undergraduate class and no graduate students. The student-to-professor ratio is 9-to-1, and the average course enrolls 17 students. Students are generally well prepared for college learning in general and for quantitative thinking in particular; on the SAT mathematics test the inter-quartile range is 660 to 760. Campus-wide attention has been paid to QR for the last eight years. In one recent, 18-month period there were more than 125 different, faculty participants at various QR professional development events; that is roughly 65% of the regular faculty. As noted in the introduction, this broad engagement has increased student QRW performance as a whole. Thus, the question explored in this paper is "Which courses have *the greatest* positive impact on QRW?" and not so much "Which courses have a positive impact on QRW?" Were an institution with a different context to carry out a similar investigation, results might differ from those seen at Carleton.

Even within the context of the broader Carleton environment, the particularities of the teacher may matter. Specifically, Professor Lutsky was the founding director of Carleton's quantitative reasoning initiative. In that role, he was central in the development of the assessment rubric. Indeed, the first-year seminar was designed to teach to the outcomes evaluated in that rubric. To test the hypothesis that it is Professor Lutsky rather than the course which is effective, I re-ran the analysis using only students who took the seminar in Fall 2007 when I also taught a section of the course. The results are no weaker in this variation, suggesting that the course rather than the instructor is responsible for the observed patterns. However, because I also was involved in the creation of the rubric it is still possible that others who attempt to teach the "same" course may not produce learning gains as large as those reported above.

Three sources of selection bias may also influence the results, particularly the findings that completing Calculus II and III, natural science courses, and the QR first-year seminar predict stronger QRW scores. The first potential bias flows from student selection of classes. Because students are not randomly assigned to courses, it is possible that students with strong quantitative skills select into the QR first-year seminar or natural science courses. In this case, all or part of the positive findings above could be driven by reverse causality. Two facts argue against this interpretation of the data. First, QRW performance is not correlated with all of the courses for which this story of selection bias would seem to be relevant. The QR community has always argued that QR is not simply mathematics by a different name and the observed lack of correlation between QRW scores and courses like Calculus I suggests we are right.

Moreover, the correlations between standardized measures of mathematical ability, course-taking, and QRW scores do not support this selection bias story. Students who registered for the first-year seminar have higher standardized mathematics test scores, but not dramatically so. When compared to the average of all other students, their average ACT and SAT mathematics scores are 0.77 and 31 points higher, respectively. When natural science courses completed is regressed on ACT or SAT mathematics scores, while the coefficients are statistically significant they are quite modest. A 1-point change in the ACT score increases the number of completed science courses by less than one-tenth. A 30-point increase in the SAT score has a similarly modest effect. Even if the standardized measures of "mathiness" were closely correlated with course-taking, the mathematics test scores do not predict QRW scores. In an ordered probit analysis available from the author on request, even large changes in the ACT mathematics score predict meager changes in both central and peripheral QRW and neither effect is statistically significant. Given that mathematics test scores neither correlate with the independent nor the dependent variable, it seems unlikely that course selection based on mathematical preparation explains the results above. Of course, the rate of selection based on mathematical ability may be different at institutions with different standardized test score profiles.

If students aren't selecting on mathematical ability, perhaps they are selecting on writing ability. Given that our measure of quantitative reasoning performance situates the practice in the context of argument, is it possible that our results reflect more of the students' writing ability than quantitative acumen? If this were the case we would expect to see students who take many writing-intensive courses score highly in the assessment. To test this idea, I replicated the analysis using the number of English courses completed as the dependent variable. In results available on request, the analysis showed that while relatively higher rates of English course-taking significantly predicts quality of central QRW, the sign of the effect is negative. Similarly, neither the ACT English nor the SAT verbal test scores predict QRW. These results suggest that the observed positive effects of upper-level calculus, natural science courses, and the first-year seminar are not simply reflecting better writing ability among student who selected those courses.

The final source of selection is in the submission of papers to the writing portfolio. For example, students who took the first-year seminar might submit papers from that class. Moreover, those papers were written in an environment designed specifically to support good QRW practice. This selection of papers might be another source of bias. To test this hypothesis, I re-analyzed the effect of QR courses excluding all observations in which the assessed paper was written for the first-year seminar. This exclusion has no meaningful impact on the results.[12]

In a similar way, students who register for science courses may submit papers from those courses. I re-analyzed the correlation between science course completion and centrally relevant QRW quality, excluding all papers written for science courses. This sample restriction is quite aggressive, affecting 112 of the 301 observations. The estimates no longer point to a significantly positive effect of science course taking. This suggests that the previously reported correlation between science course completion and central QRW proficiency may be the result of selection bias related to the choices students make in submitting portfolio papers. However, it should be noted that it is an odd examination of central QR that excludes the roughly 1/3 of such papers written in the natural sciences. It may be that this attempt to eliminate selection bias is throwing the baby out and studying the bath water.

Concluding Remarks

Recognizing the caveats discussed at the end of the previous section, the results of this study suggest that QRW can be improved through courses which place significant focus on quantitative communication—specifically, in this case, Professor Lutsky's Measured Thinking seminar and natural science courses. At the same time, little correlation was found with course completion in introductory statistics, principles of economics, social sciences, and an array of broadly distributed courses revised to include at least one QR assignment. What can be made of these correlations and non-correlations?

Dealing with the latter first, we should bear in mind two key considerations. First, as noted in the introduction, the sample of papers we assessed was taken in the context of a campus-wide initiative to improve QR, and prior research shows that general improvement has been evident. So, it is possible that students are learning better QRW in these courses even if the data say they are not learning more in these courses than in other courses. Second, the assessment tool was designed to evaluate a specific facet of QR—its use in communication. While this may be an important component of QR, it is clearly not the whole of QR. Courses which are not correlated with stronger QR in writing may markedly improve student's QR skill set (as opposed to the communication of applications of that skill set) which could be detected by other assessment instruments that

[12] Only two observations were affected, both of them involving papers with central QR relevance. One received a quality score of 2 while the other received a score of 4.

focused on demonstration of skills rather than voluntary application of those skills in an applied context.

Alternatively, it may be that these courses do teach students about QR communication, but that students require multiple treatments before they are able to transfer this skill to a new situation.[13] The literature on knowledge transfer has long shown that students are better able to apply concepts in new situations when they are taught with multiple examples (see Atkinson et al., 2000, for a review of that literature.) In fact, students who are given multiple examples do better than those given a single example along with an explicit and detailed description of the relevant procedure. It appears that having at least two examples helps students distinguish between features that are superficial from what Atkinson et al. term "deep structure." This aspect of learning theory could explain why students who took many science courses or the QR FYS (which, because it was focused on QR, could devote time to multiple examples) showed greater QRW proficiency than those who took a single principles of economics or math/statistics course. It may be that these students saw what amounts to a single example of the relevant QR principles and that subsequent course-taking is required to unlock their potential to exhibit latent learning gains.

Given how easy it is to find explanations for how these courses which are not correlated with stronger QRW scores are nonetheless building student QR capacity, it would be inappropriate to conclude based on these results alone that these courses do not meet QR objectives. In a very real sense, the lack of correlation between QRW performance and some courses points to the need for further study.

By contrast, the observed positive correlations between QRW and upper-level calculus, natural science courses, and a QR FYS seem to signal that intentional QR teaching can improve student outcomes. The results from the dedicated QR seminar appear to be clearest with large effect sizes and less question about selection bias. However, like James Madison University (2012), I find evidence that natural science courses also increase student performance particularly in central QR (though this result may be confounded by paper selection bias). This finding is also in line with Meisels (2010 p. 1) who argues, citing a definition from Wikipedia, that "the core objective of science literacy is almost identical to that of numeracy 'to reason with numbers and other mathematical concepts' and 'to be comfortable with logic and reasoning.'" The fact that completion of science courses predicts better QRW performance suggests that it is possible for a large number of courses taught across many departments to support QR goals. But reflecting on the nature of the curricula in these courses, it

[13] I thank Milo Schield for this observation.

would also seem that success may depend on addressing quantitative methods and QR communication in a deep, persistent, recurring, authentic way.

Another characteristic shared by the courses which predicted the strongest QRW growth is their combination of writing and quantitative analysis. Madison (2012) argues that this combination is much more than the sum of its parts and so should be taught together. Surely, the results of this study are insufficient to point to any one factor as a definitive elixir for improving QRW, but it may prove a jumping-off point to consider these and many other possibilities as we attempt to create a more effective QR curriculum.

Acknowledgments

This paper was substantially improved by comments from Neil Lutsky, Milo Schield, and participants at the 2012 National Numeracy Network annual meeting. Funding was in part from the Fund for the Improvement of Post-Secondary Education (P116B040816), the National Science Foundation (DUE-0717604), and the WM Keck Foundation.

References

Atkinson, R. K., S. J. Derry, A. Renkl, and D. Wortham, 2000. Learning from examples: Instructional principles from the worked examples research, *Review of Educational Research* 70 (2): 181–214. http://dx.doi.org/10.3102/00346543070002181

Boersma, S., and D. Klyve. 2013. Measuring habits of mind: Toward a prompt-less instrument for assessing quantitative literacy. *Numeracy* 6 (1): Article 6. http://dx.doi.org/10.5038/1936-4660.6.1.6

Boersma, S., C. Diefenderfer, S. W. Dingman, and B. L. Madison. 2011. Quantitative reasoning in the contemporary world, 3: Assessing student learning. *Numeracy* 4 (2): Article 8. http://dx.doi.org/10.5038/1936-4660.4.2.8

Bressoud, D. 2009. Establishing the quantitative thinking program at Macalester. *Numeracy* 2 (1): Article 3. http://dx.doi.org/10.5038/1936-4660.2.1.3

Grawe, N. D. 2011. Beyond math skills: Measuring quantitative reasoning in context. In *Measuring complex general education student learning outcomes* ed. J. D. Penn, *New Directions in Institutional Research, Special Issue: Assessing Complex General Education Student Learning Outcomes* 2011 (149): 41–52. http://dx.doi.org/10.1002/ir.379

———. 2012. Achieving a quantitatively literate citizenry: Resources and community to support national change. *Liberal Education* 98 (2): 30–35.

———, N. S. Lutsky, and C. J. Tassava. 2010. A rubric for assessing quantitative reasoning in written arguments. *Numeracy* 3 (1): Article 3. http://dx.doi.org/10.5038/1936-4660.3.1.3

Hillyard, C. 2012. Comparative study of the numeracy education and writing across the curriculum movements: Ideas for future growth. *Numeracy* 5 (2): Article 2. http://dx.doi.org/10.5038/1936-4660.5.2.2

James Madison University. 2012. General education assessment reports, Cluster Three. http://www.jmu.edu/assessment/JMUAssess/GenEdOverview.htm (accessed October 2, 2012).

Madison, B. L. 2012. If only math majors could write…. *Numeracy* 5 (1): Article 6. http://dx.doi.org/10.5038/1936-4660.5.1.6

Meisels, G. G. 2010. Science Literacy: Hand in glove with numeracy. *Numeracy* 3 (2): Article 1. http://dx.doi.org/10.5038/1936-4660.3.2.1

Steen, L. A. 2004. *Achieving quantitative literacy: An urgent challenge for higher education*. Washington, DC: Mathematical Association of America.