

2011

Quantitative Reasoning in the Contemporary World, 3: Assessing Student Learning

Stuart Boersma

Central Washington University, boersmas@cwu.edu

Caren Diefenderfer

Hollins University, cdiefenderfer@hollins.edu

Shannon W. Dingman

University of Arkansas, sdingman@uark.edu

Bernard L. Madison

University of Arkansas, bmadison@uark.edu

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

 Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Boersma, Stuart; Diefenderfer, Caren; Dingman, Shannon W.; and Madison, Bernard L. (2011) "Quantitative Reasoning in the Contemporary World, 3: Assessing Student Learning," *Numeracy*: Vol. 4 : Iss. 2 , Article 8.

DOI: <http://dx.doi.org/10.5038/1936-4660.4.2.8>

Available at: <http://scholarcommons.usf.edu/numeracy/vol4/iss2/art8>

Quantitative Reasoning in the Contemporary World, 3: Assessing Student Learning

Abstract

In this third paper in a series describing the Quantitative Reasoning in the Contemporary World course, the authors provide an adaptation of the Association of American Colleges and Universities quantitative literacy VALUE rubric. Describing achievement levels in six core competencies (interpretation, representation, calculation, analysis/synthesis, and communication), the resulting Quantitative Literacy Assessment Rubric (QLAR) is applicable to grading student work and has exhibited a high degree of reliability in two separate scoring tests (97% and 88% respectively). The distribution of the six core competencies across the 24 case studies in the authors' quantitative reasoning casebook shows that interpretation, calculation, and analysis/synthesis were present in most all of the case studies. In addition to acting as a reliable scoring tool, the QLAR can improve teaching, learning, and curricular materials.

Keywords

quantitative literacy, quantitative reasoning, assessment rubric, student learning

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

Stuart Boersma is professor and former Chair of the Department of Mathematics, Central Washington University. He enjoys writing expository mathematics papers and received the 2005 Trevor Evans Award of the Mathematics Association of America.

Caren Diefenderfer is professor of mathematics at Hollins University where she has chaired the mathematics and statistics department as well as the division of natural and mathematical sciences. She served as Chief Reader for AP Calculus from 2004-2007.

Shannon Dingman is an assistant professor in the Department of Mathematics at the University of Arkansas. His research specialty is mathematics education, including such topics as the impact of policy on mathematics education.

Bernard L. Madison is professor and former Chair of the Department of Mathematics, University of Arkansas, and former Dean of its Fulbright College of Arts and Sciences. He was founding president of the National Numeracy Network and is a frequent contributor to this journal.

Introduction

In the decade since quantitative literacy (QL) was described in *Mathematics and Democracy* (Steen 2001), QL, or numeracy, has gained more prominent attention and become better understood. At the same time, the scope of QL has become broader than earlier conceptions, reaching across multiple disciplines and outside of traditional areas of learning and assessment (Madison and Steen 2008a, b). Assessing QL has always presented challenges, and several writers have noted the hurdles (e.g., Wiggins 2003; Shavelson 2008). Additionally, several institutions (e.g., Hollins University, James Madison University, Trinity College, Virginia Commonwealth University, Wellesley College) have created instruments and processes to assess QL or quantitative reasoning (QR) as an institutional learning goal. Most items on QL assessment instruments have been multiple-choice or short answer to facilitate machine scoring. Scoring more complex answers requires the development of scoring rubrics that can be applied consistently by multiple readers. One such rubric has been developed for scoring quantitative argument in writing at Carleton College (Grawe et al. 2010), while another was created at Macalester College to measure longitudinal growth in QL.¹ In this paper we report on the process of adapting a rubric that was originally designed for assessing QL or QR at the institutional level to create a rubric for scoring student work in a QR course. We present the modified rubric, give results of mapping study questions from QR case studies to the core competencies, and report on the consistency of scoring by different readers. Finally, we give instances showing why the rubric would assist in improving teaching, learning, and curricular materials.

Over the past six years, interrelated courses in QR have evolved at the University of Arkansas, Central Washington University, and Hollins University. In a previous issue of *Numeracy*, the Arkansas course and its challenges were described (Dingman and Madison 2010) and observations and questions about how students learn to reason quantitatively were presented (Madison and Dingman 2010). This paper is the third component of the results of research and development surrounding these courses.

From 2007 to 2011, expansion of these courses was supported by the National Science Foundation (DUE 0715039), with the authors as Principal Investigators. The courses have become identified with the title of the NSF project, Quantitative Reasoning in the Contemporary World (QRCW).² One of the products of this effort was the creation of a book of 24 case studies of news media articles, *Case Studies for Quantitative Reasoning: A Casebook of Media*

¹ http://serc.carleton.edu/quirk/pkal_workshop10/assessment_tools.html

² Project website: <http://www.cwu.edu/~boersmas/QRCW/>

Articles, Second Edition (Madison et al. 2009), which we will refer to as the QRCW casebook. Each case study is centered on one or more artifacts from public media, mostly newspaper and magazine articles. Each case study focuses on the quantitative content of the articles, and the core QR challenges consist of study questions about the quantitative content of the article. One of the case studies can be found in Appendix A to Dingman and Madison (2010).³ Although each case study has some warm-up exercises that scaffold the students' thinking to the study questions, the study questions are the proof of the pudding.

The most challenging aspect of the QRCW course has been, as expected, assessment. In fact, Dingman and Madison (2010) described assessment as one of the major challenges in developing the QRCW course at Arkansas. The initial efforts at assessment were directed toward evaluation of the course. For three semesters at Arkansas, one semester at Hollins, and one quarter at Central Washington in 2007–2008, a pre- and post-test and an attitude survey were administered to all students enrolled in the QRCW course. Some of the results are reported in Dingman and Madison (2010). From the beginning of the development of the QRCW courses, learning goals have been stated, sorted in various ways, and mapped to the content of the case studies. Various options for assessing student work were considered. One was development of a multiple-choice test that would serve as a reliable proxy for a more in-depth assessment of students' responses to free-response essay questions. Another option became available when the Association of American Colleges and Universities (AAC&U) published its rubric for assessing undergraduate student learning in QL at the end of 2009. The creation of this rubric provided the opportunity to conform our thinking to this national effort. QL was one of fifteen areas of student learning in the AAC&U project titled Valid Assessment of Learning in Undergraduate Education (VALUE). As stated in the AAC&U Web site,⁴

The VALUE project aimed at rubrics to position learning at all undergraduate levels within a basic framework of expectations such that evidence of learning can be shared nationally through a common dialog and understanding of student success. The core expectations articulated in all 15 of the VALUE rubrics can and should be translated into the language of individual campuses, disciplines, and even courses.

This paper describes our translation of the QL VALUE rubric to a QL Assessment Rubric (QLAR) in order to assess student work in QRCW courses.

Implications beyond the QRCW Materials and Courses

In this paper we will address the importance of using a specific QLAR in developing specific QL course materials. This approach would improve the

³See <http://services.bepress.com/numeracy/vol3/iss2/art4/>

⁴http://www.aacu.org/value/rubrics/index_p.cfm?CFID=1814627&CFTOKEN=33120747

organization and wording of study questions in the QRCW casebook as well as other efforts. Also, the core competencies that comprise the areas of scoring student work provide a structure for helping to organize how students learn to reason quantitatively. In fact, these two outcomes are likely more important than the actual scoring of student work. Whatever curricular materials are used, these two outcomes apply. Since the core competencies in this study are taken from the national effort by AAC&U, they provide an anchor for further studies using other QR materials and courses. Of course, complex student-produced answers to QR questions that elicit reasoning more than content and methodologies are more difficult to assess, and so the use of the rubric in this study provides a model that will apply beyond the curricular materials used.

Part of this study involved mapping the study questions to the core competencies, i.e., identifying which of the core competencies are needed to answer the study questions. We give the proportions of the questions that require each competency. For example, approximately two-thirds of the questions require interpretation. This raises the question of what would be the analogous proportions for other QR materials, particularly textbooks that are more traditional than the QRCW casebook. A survey of other materials would be of interest, but that is beyond the scope of the current study.

The AAC&U VALUE Project

In late 2009, AAC&U published the results of its VALUE project (AAC&U 2009). A major product of this project was the development of rubrics for institution-level use in evaluating and discussing student learning (i.e., they were not intended for grading student work). Rubrics were created for fifteen areas, ten of which are termed intellectual and practical skills. The ten intellectual and

Table 1

Definition of Quantitative Literacy, Critical Thinking and Information Literacy Adopted by the VALUE Project (AAC&U 2009)

<u>Quantitative Literacy</u> (QL) – also known as Numeracy or Quantitative Reasoning (QR)	a "habit of mind," competency, and comfort in working with numerical data. Individuals with strong QL skills possess the ability to reason and solve quantitative problems from a wide array of authentic contexts and everyday life situations. They understand and can create sophisticated arguments supported by quantitative evidence and they can clearly communicate those arguments in a variety of formats (using words, tables, graphs, mathematical equations, etc., as appropriate).
<u>Critical thinking</u>	a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.
<u>Information Literacy</u>	the ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively and responsibly use and share that information for the problem at hand.

practical skills are: inquiry and analysis, critical thinking, creative thinking, written communication, oral communication, reading, quantitative literacy, information literacy, teamwork, and problem solving. Four are termed personal and social responsibility: civic knowledge and engagement, intercultural knowledge and competence, ethical reasoning, and foundations and skills for lifelong learning. The final area is integrative and applied learning.

The areas of student learning overlap. For example, critical thinking, information literacy and quantitative literacy are three of the ten intellectual and practical skills and these three have significant overlap. Some authors, for example Bok (2006), include QL as a part of critical thinking. Each of these three areas is defined in Table 1 as in the VALUE rubrics.⁵

The core competencies are given for each area with four levels of achievement. The six competencies for QL are interpretation, representation, calculation, application/analysis, assumptions, and communication. Critical thinking has five core competencies: explanation of issue, evidence (selecting and using information to investigate a point of view or conclusion), influence of context and assumption, student's position (perspective, thesis/hypothesis), and conclusions and related outcomes (implications and consequences).

The QLAR

The QL Assessment Rubric (QLAR) is an adaptation of the AAC&U VALUE QL rubric to make it more applicable to grading student work. The matrix of competencies and achievement levels is given in Figure 1. The QLAR is intended to measure achievement levels of the associated QL core competencies in a variety of assignments. Depending on the nature of the prompt, not all competencies will be present in every student response. Occasionally there may be a perceived overlap of the QL core competencies, and in this case one competency should be chosen as the dominant one to assess. Careful wording of the prompt is often crucial to obtaining a focused student response. These issues will be discussed in more detail later in this paper.

To develop the QLAR, we have slightly modified the core competencies (names and definitions) and have rewritten the milestone descriptors that we call achievement levels. Each of the case studies in the QRCW casebook begins with an account of the learning goals of the case study. These reflect our thoughts about the core competencies that are necessary to answer the specific study questions. Two of those core competencies were not immediately apparent in the six competencies in the QL VALUE rubric. Our missing competencies were critical reading and number (or quantity) sense. Initially, we considered adding these two competencies, but in order to be aligned with the national AAC&U

⁵ http://www.aacu.org/value/rubrics/index_p.cfm?CFID=1814627&CFTOKEN=33120747

Quantitative Literacy Core Competency	Achievement Level			
	3	2	1	0
Interpretation <i>Ability to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)</i>	Correctly identifies all relevant information.	Correctly identifies some, but not all, relevant information.	Some relevant information is identified, but none is correct.	No relevant information identified.
Representation <i>Ability to convert information from one mathematical form (e.g. equations, graphs, diagrams, tables, words) into another.</i>	All relevant conversions are present and correct.	Some correct and relevant conversions are present but some conversions are incorrect or not present.	Some information is converted, but it is irrelevant or incorrect.	No conversion is attempted.
Calculation <i>Ability to perform arithmetical and mathematical calculations.</i>	Calculations related to the problem are correct and lead to a successful completion of the problem.	Calculations related to the problem are attempted but either contain errors or are not complete enough to solve the problem.	Calculations related to the problem are attempted but contain errors and are not complete enough to solve the problem.	Calculations given are not related to the problem, or no work is present.
Analysis/Synthesis <i>Ability to make and draw conclusions based on quantitative analysis.</i>	Uses correct and complete quantitative analysis to make relevant and correct conclusions.	Quantitative analysis is given to support a relevant conclusion but it is either only partially correct or partially complete (e.g. there are logical errors or unsubstantiated claims).	An incorrect quantitative analysis is given to support a conclusion.	Either no reasonable conclusion is made or, if present, is not based on quantitative analysis.
Assumptions <i>Ability to make and evaluate important assumptions in estimation, modeling, and data analysis.</i>	All assumptions needed are present and justified when necessary.	At least one correct and relevant assumption is given (perhaps coupled with erroneous assumptions), yet some important assumptions are not present.	Attempts to describe assumptions, but none of the assumptions described are relevant.	No assumptions present.
Communication <i>Ability to explain thoughts and processes in terms of what evidence is used, how it is organized, presented, and contextualized.</i>	A correct and complete explanation is clearly presented.	A partially correct relevant explanation is present, but incomplete or poorly presented.	A relevant explanation is present, but is illogical, incorrect, illegible, or incoherent.	No relevant explanation is provided.

Figure 1. the QLAR.

effort, we decided that we could include these missing competencies under the AAC&U interpretation competency. Some of the changes we made to the VALUE rubric to produce the QLAR are as follows:

1. Included a column for a score of 0. The VALUE rubric had a default score of zero if the work did not meet the level-one benchmark, and a score of zero was also assigned if the core competency was not present as a part of the answer. We include a zero column to more clearly acknowledge the presence or absence of core competencies..
2. Removed the column for a score of 4. In the VALUE rubric the score of four designated a capstone achievement. Capstone proficiency requires a cumulative experience over a complete undergraduate curriculum and is not a normal consideration for student work on a specific assignment in a stand-alone QL course.
3. Provided more objective descriptions of achievement levels. Several of the descriptions in the VALUE rubric made distinctions between levels difficult because of the use of qualitative words. For example, distinguishing between “workmanlike” and “competent” or between “plausible” and “reasonable” was found to be too subjective.
4. Changed the core competency of application/analysis to analysis/synthesis. This change was made to accommodate drawing conclusions by either analyzing (that is, breaking apart circumstances) or synthesizing (that is, pulling together components).

The definition of QL that we use in the QRCW project and courses is a version of the mission statement of the National Numeracy Network,⁶ which draws heavily on the description of enabling literacy by Cremin (1988) as quoted by Orrill (2001). The quantitative literacy we seek to assess includes command of both the enabling skills needed to search out quantitative information and the power of mind necessary to critique, reflect upon, and apply the quantitative information in making decisions. This definition is consistent with the more detailed definition of QL in the VALUE rubric.

As we have applied the QLAR to student responses to prompts (study questions) in the QRCW casebook we have encountered questions concerning both the mapping of prompts to core competencies and what scores to assign to responses. However, as is detailed below, we have achieved an overall high level of agreement in both mapping of prompts and score levels. Because our mapping and scoring data are based on the QLAR as presented here, we resisted making any further refinements based on our findings. Nonetheless, refinements and

⁶ See <http://serc.carleton.edu/nnn/index.html>

changes in the QLAR likely will make it more effective. We plan to continue to refine the QLAR as we use it and encourage others to do likewise.

Mapping the QRCW Case Study Questions to the QLAR Core Competencies

We carefully examined every study question in the QRCW casebook to determine which core competencies were necessary for a complete solution. For example, case study 4.3 (Figure 2)⁷ in the QRCW casebook highlights an article by Jerry Taylor (2001) with the headline, “Forcing fuel efficiency on consumers doesn’t

1. Locate the following quantitative assertions in the article. Are there others? What are they?

The quantitative assertions include the following:

 - I. Economists have discovered that, over the long run, a 20 percent increase in gasoline costs, for instance, will result in a 20 percent decline in gasoline consumption.
 - II. A recent report from the National Academy of Sciences, for instance, notes that the fuel efficiency of a large pickup could be increased from 18.1 miles per gallon to 26.7 miles per gallon at a cost to automakers of \$1,466.
 - III. But do the math: It would take the typical driver 14 years before he would save enough in gasoline costs to pay for the mandated up-front expenditure [\$1466].
 - IV. A similar calculation for getting a large SUV up to 25.1 miles per gallon leads to a \$1,348 expenditure and, similarly, more than a decade before buyers would break even.
 - V. You could take that \$1,466, for instance, put it in a checking account yielding 5 percent interest, and make a heck of a lot more money than you could by investing it in automobile fuel efficiency.

2. Answer the following:
 - a. Which of the assertions can be checked without considerable research?
 - b. What assumptions would need to be made in checking assertion III?
 - c. What assumptions would need to be made in checking assertion IV?
 - d. What assumptions would need to be made in checking assertion V?

3. Answer the following:
 - a. Is the assertion III above reasonable? Explain why or why not.
 - b. What would be the effect of increased costs of gasoline on assertion III?
 - c. What would be the effect of increased miles driven per year on assertion III?
 - d. Assume the cost of gasoline in 2001 was \$1.40 per gallon and that it would take 14 years for the “typical driver” to recover the \$1466 through savings in gasoline costs. How many miles per year would the “typical driver” drive?

Figure 2. Study Questions from Case Study 4.3 (Madison et al 2009).

⁷ See also appendix A in Dingman and Madison (2010).

work.” In the article, the following is asserted: “It would take the typical driver 14 years before he could save enough in gasoline costs to pay for the \$1466 required to modify a motor vehicle so that the fuel efficiency changes from 18.1 miles per gallon to 26.7 miles per gallon” (Assertion III). One study question (2b) asks what assumptions one must make in order to check this assertion. The only core competency here is assumptions, so this study question would map to one competency, assumptions. Another study question (3a) asks, “Is this assertion reasonable? Explain why or why not.” This requires representation (generating a cost equation), calculation, analysis/synthesis (drawing a conclusion), and communication (explaining), so this question would map to these four competencies.

The 24 case studies in the casebook have a total of 234 study questions. Each study question was mapped to one or more of the core competencies in the QLAR. Two of us mapped the study questions in the first half of the casebook, and two others mapped the second half. Following that, we swapped study questions to determine if each pair would independently create the same mapping as the other. Agreement of the mapping was substantial and all differences were resolved in discussions. Table 2 shows the number of case studies (24), the number of study questions (234), the number of competencies represented in answering the questions (up to 6 for each question), and the number that we agreed to on the first iteration of the process using the modified rubric.

Table 2
Agreement of Core Competencies in QRCW Casebook

Case studies	Study questions	Competencies	Agreement
24	234	467	437 (94%)

One of the points of disagreement between the two groups of coders was whether a question required communication or analysis/synthesis. Obviously, most answers require some communication, but communication in this rubric is “*Explaining thoughts and processes in terms of what evidence is used, how it is organized, presented, and contextualized.*” Analysis/synthesis is “*Making and drawing conclusions based on quantitative analysis.*” Drawing a distinction is sometimes difficult, but scoring under one or the other is usually the resolution. The core competency of communication, as we applied it, concerns the explanation of a process, that is, a description of the thinking and how conclusions were obtained. The actual thinking and conclusions are part of the analysis/synthesis competency.

Another point of disagreement concerned the overlap of interpretation and communication, since both could involve explanations. Our final definition of interpretation allows for interpreting explanations to be scored under the inter-

pretation competency: Interpretation—*Ability to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)*. Therefore explaining mathematical information, whether it was gleaned from text by students or given in the question prompt, was assigned to interpretation and not communication. The competency of communication focuses on the explanation as described in the above paragraph.

The distribution of the core competencies across the 24 case studies and the 234 study questions reveals how the competencies are addressed in the casebook. Table 3 shows the frequency and prevalence of the six core competencies in the casebook. Additionally, all six competencies were required in three case studies, five were required in 13 case studies, four in seven, and one case study required only two.

Table 3
Frequency and Prevalence of Core Competencies in the QRCW Casebook

Competency	Number of questions	Percent of questions	Number of case studies ($N=24$) which have at least one question mapped to competency	Number of case studies which have 50% or more questions mapped to competency
Interpretation	152	65	24	18
Representation	68	29	19	6
Calculation	101	43	21	9
Analysis/Synthesis	79	34	20	7
Assumptions	14	6	10	0
Communication	53	23	18	4

The above distribution is not presented as a model. In fact, the casebook was compiled without these core competencies or any other set of competencies as a guide. The only effect of core competencies or learning goals guiding the development of the casebook was in the intuition and experience of the authors. However, a set of competencies such as the six here could serve as a guide for developing study materials for QL. In fact, in looking at the overall breakdown of competencies across the case studies, we notice that we did not create many questions that required students to evaluate or make assumptions in order to solve problems. If, for example, early in a course one wishes to emphasize interpretation, representation, and calculation, case studies could be so selected. Later in the course the more complex competencies of analysis/synthesis, assumptions, and communication could be incorporated in the selection of case studies and the creation of specific prompts.

Estimating Reliability

As mentioned above, one of the major changes we made to the AAC&U QL VALUE rubric was the rewriting of the milestone, or achievement level, descriptors. When we attempted to use the QL VALUE rubric to score individual student work, we could not produce very reliable results. We decided this was due to the use of qualitative words in the milestone descriptors. For example, to achieve a score of 3 under the application/analysis competency, a student must use "... quantitative analysis of data as the basis for *competent* judgments, drawing *reasonable* and appropriately qualified conclusions from this work" (emphasis added). A score of 2 would be warranted if the student used "... quantitative analysis of data as the basis for *workmanlike* (without inspiration or nuance, ordinary) judgments, drawing *plausible* conclusions from this work" (emphasis added). We found it difficult to make distinctions between competent judgments and workmanlike judgments. Similarly, different users of the rubric often had differing views on the difference between a reasonable conclusion and a plausible conclusion. As mentioned above, the difficulties we encountered in interpreting some of the milestone descriptors led us to develop the QLAR. In order to test the reliability of the QLAR we designed a simple reliability study.

Our scoring reliability study in applying the QLAR concentrated on questions 2 (b,c,d) and 3 (a,b,c,d) concerning the "Forcing fuel efficiency" case study discussed previously ((Fig. 2; see also Dingman and Madison 2010, Appendix A). These two questions were dealing with the assertions in question 1, which is also shown in Figure 2.

Students in two sections of one of the QRCW courses produced answers to seven of these eight questions in 2 and 3. Question 2a was not assigned because of the subjective nature of possible answers. This was an assignment to be graded by the QRCW course instructor, so the students were motivated to produce creditable answers. We had previously agreed on which core competencies were required in each of the answers. These are listed in the Table 4.

Table 4
Core Competency for Study Questions

Question	Competencies
2b	Assumptions
2c	Assumptions
2d	Assumptions
3a	Representation, calculation, analysis/synthesis, communication
3b	Interpretation, analysis/synthesis
3c	Interpretation, analysis/synthesis
3d	Representation, calculation

These student papers were used to conduct two independent reliability tests of applying the QLAR. Two of us (Readers A and B) took a sample of the student papers from one section, and the other two (Readers C and D) took a sample of papers from the other section. Readers A and B scored 3a, 3b, 3c, and 3d on 24 papers. Readers C and D scored 2b, 2c, 2d, 3a, and 3d on 14 papers. In discussions before the scoring-reliability exercise, we recognized the need for “reader training.” This involved making sure we understood what answers were expected and what parts of those answers were likely to be scored under various core competencies. Sample answers to the study questions are posted on a password-protected portion of the QRCW Web site⁸ and we reviewed those to better understand what we expected in an optimal response. We discovered that a few of our sample answers on the site would not have received top scores. We learned the following valuable lesson: knowing the rubric for scoring answers is important in the development and publishing sample answers that are supposed to be models.

Using only the descriptors present in the QL Assessment Rubric, Readers A and B independently scored the achievement level of each competency for each question. This led to 238 assessment scores per reader as shown in Table 5.

Table 5
Number of Assessment Scores Per Core Competency

QL Competency	Number of Assessment scores per reader
Interpretation	48
Representation	48
Calculation	46
Analysis/Synthesis	72
Communication	24
TOTAL	238

Data for first reliability test with Readers A and B

Of the 238 assessment scores, Readers A and B disagreed on seven, or 2.94%, and agreed on 231, or 97.06%. When Reader A and B disagreed, the average difference in their assessment scores was 1.23. Ironically, A’s and B’s assessment scores showed the greatest level of disagreement on the QL competency “calculation.” The disagreement level for Readers A and B on all five core competencies is shown in Table 6.

The second reliability test was conducted to measure the reliability of the QLAR with questions 2b, 2c, 2d, 3a, and 3d. Student work ($N=14$) from a different QRCW section was collected and distributed to two different readers (Reader C and Reader D). After a brief discussion of the rubric, Readers C and D

⁸ <http://www.cwu.edu/~boersmas/QRCW/>

independently scored the achievement level of each competency for each question. This led to 126 assessment scores per reader as shown in Table 7.

Table 6
Disagreement Level Per Core Competency

QL Competency	Number of disagreements	<i>N</i>	disagreement
Interpretation	1	48	2.08
Representation	1	48	2.08
Calculation	3	46	6.25
Analysis/Synthesis	2	72	2.78
Communication	0	24	0

Data for first reliability test with Readers A and B

Table 7
Number of Assessment Scores Per Core Competency

QL Competency	Number of Assessment scores per reader
Assumptions	42
Representation	28
Calculation	28
Analysis/Synthesis	14
Communication	14
TOTAL	126

Data for second reliability test with Readers C and D

Of the 126 assessment scores, Readers C and D disagreed on 15, or 11.9%, and agreed on 111, or 88%. When Readers C and D disagreed, the discrepancy was almost always 1. Additionally, C's and D's assessment scores showed the greatest level of disagreement on the QL competencies of "communication" and "calculation." Many of the discrepancies within the calculation competency concerned how to score work where student calculations were correct, yet the student provided incorrect units for the computed answer. The disagreement level for Readers C and D on all five core competencies is shown in Table 8.

Table 8
Disagreement Level Per Core Competency

QL Competency	Number of disagreements	<i>N</i>	% disagreement
Assumptions	2	42	4.76
Representation	2	28	10.71
Calculation	6	28	21.43
Analysis/Synthesis	0	14	0
Communication	5	14	35.7

Data for second reliability test with Readers C and D

These two reliability exercises indicate that the QLAR can be used to consistently score student work and that consistent scoring can result from multiple readers. The experience does indicate that careful analysis of possible student answers before scoring is necessary for consistency. Scoring a sample of student responses and then discussing the results can reveal issues that will occur in the scoring. For example, how do incorrect units affect scores, and in which competencies does one wish to score the student work?

Discussion

Scoring of Student Papers for Course Grades

Although the scoring exercise described above was applied to student work that was an assignment for part of a course grade, the scoring for the course grade was independent of the results of the reliability exercise for several reasons. First, the students were not familiar with the QLAR rubric. Further, the QLAR produced scores that were not normalized by any weighting due to the level of difficulty of the study questions. For example, the highest possible QLAR score for 3a (Fig. 2) is 12, four scores of 3, while the highest possible QLAR score for 3d is 6. A response yielding 12 on 3a can be relatively simple, while the representation and calculation required for 3d is more challenging. However, on a sample of 14 papers where we knew both the QLAR score and the instructor-produced score with no normalizing or weighting, the correlation was 0.78.

Scoring of papers for course grades in QRCW courses is accomplished by individual instructors, guided by discussions among the instructors and the sample answers posted on the QRCW Web site. This keeps the grading of students in the courses a matter between the instructor and students; in some cases this is rigorous and creditable and less so in others. Using a rubric such as the QLAR with appropriate preparation and weighting of scores would produce creditable scores, but much of the value of a rubric such as the QLAR lies in other areas as described below: guiding student thinking and production of instructional materials.

Rubric as Model for Guiding Student Thought Processes

Having a rubric in mind or overtly presented when prompting students—whether in daily study questions, in class discussions, or on examinations—allows one to align the prompts with an organized thought process as represented by the rubric. For example, answering question 3a (Fig. 2) can be partitioned into five parts: assumptions, representation, calculation, analysis/synthesis, and communication. The assumptions part was addressed in 2b. Representation is “*converting information*” about the yearly savings in gasoline costs into a cost equation that gives the costs for any number of years or a quotient of the \$1466 and the annual

savings. Then the student calculates, either the savings for 14 years or the number of years required to recover \$1466. Based on the representation and calculation, the student *“makes and draws a conclusion.”* The answer is completed via communication by *“explaining thoughts and processes in terms of what evidence is used.”*

Rubric as Guide for Development of Instructional Materials

As we mapped the study questions from the QRCW casebook to the six core competencies in the QLAR we discovered study questions that could be significantly improved by providing more explicit directions in order to elicit student responses that were better aligned with the competencies. This was particularly helpful if the question mapped to several competencies. For example, one of our prompts was, “Find out how the Standard & Poor’s 500 stock index is computed.” If we want to score that response on communication, as we did, then the prompt would be better stated as, “Describe how the Standard & Poor’s 500 stock index is computed.” Asking for an explanation will highlight the need for communication. As a second example, consider question 2a of Figure 2. Typical student responses do not allow for easy scoring by the QLAR. Simply adding the statement “Explain your reasoning” would allow one to assess, for example, interpretation. Several study questions from the casebook could be similarly improved by carefully rewriting the prompt in order to elicit student responses which contain assessable quantitative reasoning content.

The QRCW experience has pointed to the need to scaffold student thinking and performance. One significant addition to the second edition of the QRCW casebook was adding warm-up exercises to each case study. Students were not prepared for the open-ended study questions without the preliminary warm-up exercises. The same is true for many study questions that require four, five, or six core competencies. For example, sorting case studies so that earlier ones require interpretation and representation allows students to build on this understanding when more complex prompts appear later.

Conclusion

The foregoing describes how a rubric adapted from the AAC&U QL VALUE rubric was applied after the fact to the study questions in the QRCW casebook. Two major results of this application are evident. First, the model answers for the 234 study questions in the 24 case studies in the casebook map very nicely to the six core competencies, slightly modified from the AAC&U rubric. Further, each of the core competencies is required in a significant number of the study questions, with the competency of assumptions being the least required. As one would expect in case studies of media articles, interpretation is the most required

competency (2/3 of the QRCW casebook questions). Interpretation is followed in frequency by calculation at nearly 1/2, representation and analysis/synthesis at about 1/3, and communication at 1/4. As noted earlier in implications of this study beyond QRCW, what these proportions are for other QR/QL materials would be of interest.

The second major result is that the modified rubric can be applied consistently by multiple readers to score student work. With sufficient attention to understanding the desired responses to prompts, and with careful assignment of various aspects of responses to core competencies, very consistent scoring from multiple readers can be achieved.

A side result, at least a result not anticipated at the outset of this study, is the value of the modified rubric for guiding student thinking and the development of instructional materials. In fact, since applying the rubric to score student work is labor intensive, and, in many places, grading in courses is a matter between instructor and student, the real value of the modified rubric may lie in these two areas.

We hope that the creation of this rubric provides a valuable tool to the field with regards to assessment in the QL domain. We realize that further refinement of the QLAR is needed, and we urge those who use the rubric to share their thoughts and alterations with the authors and the QR community. The challenge in assessing student work will long persist; however, we feel that the ongoing dialogue and sharing of assessment tools can lessen these challenges by clearly articulating the goals and ideals that are important in enhancing students' overall abilities to reason quantitatively.

Acknowledgments

Partial funding of this work was provided by the National Science Foundation (DUE 0715039). Additional support was gratefully provided by the University of Arkansas, Central Washington University, and Hollins University. The comments of the reviewers and editor of this paper led to improvements.

References

- AAC&U. See Association of American Colleges and Universities.
 Association of American Colleges and Universities. 2009. Assessing learning outcomes: Lessons from AAC&U's VALUE project. *Peer Review*, Winter 2009, 11 (1).
http://www.aacu.org/peerreview/pr-wi09/pr-wi09_index.cfm (accessed June 7, 2011).
 Bok, D. 2006. *Our Underachieving Colleges*. Princeton, NJ: Princeton University Press.
 Cremin, L. A. 1988. *American Education: The Metropolitan Experience 1876-1980*. New York, NY: Harper & Row.

- Dingman, S. W. and B.L. Madison. 2010. Quantitative reasoning in the contemporary world, 1: The course and its challenges. *Numeracy* 3 (2): Article 4. <http://dx.doi.org/10.5038/1936-4660.3.2.4> (accessed February 14, 2011).
- Grawe, N., 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> (accessed May 25, 2011)
- Madison, B. L., S. Boersma, C.L. Diefenderfer, and S.W. Dingman. 2009. *Case studies for quantitative reasoning: A casebook of media articles*, 2nd ed. New York, NY: Pearson.
- Madison, B. L., and S.W. Dingman. 2010. Quantitative reasoning in the contemporary world, 2: Focus questions for the numeracy community. *Numeracy*, 3 (2). Article 5. <http://dx.doi.org/10.5038/1936-4660.3.2.5> (accessed February 14, 2011).
- Madison, B. L., and L.A. Steen. 2003. *Quantitative literacy: Why numeracy matters for schools and colleges*. Princeton, NJ: National Council on Education and the Disciplines. <http://www.maa.org/ql/qltoc.html> (accessed May 28, 2010).
- . 2008a. Evolution of numeracy and the National Numeracy Network. *Numeracy*, 1 (1), Article 2. <http://dx.doi.org/10.5038/1936-4660.1.1.2> (accessed February 13, 2011).
- , eds. 2008b. *Calculation vs. context: Quantitative literacy and its implications for teacher education*. Washington, DC: Mathematical Association of America. <http://www.maa.org/ql/calcvcontext.html> (accessed May 25, 2011).
- Orrill, R. 2001. Mathematics, Numeracy, and Democracy. In *Mathematics and democracy: The case for quantitative literacy*, ed L.A. Steen, xiii-xx, Princeton, NJ: National Council on Education and the Disciplines. <http://www.maa.org/ql/fm13-20.pdf> (accessed May 25, 2011).
- Shavelson, R. J. 2008. Reflections on quantitative reasoning: An assessment perspective. In *Calculation vs context: Quantitative literacy and its implications for teacher education*, ed. B. L. Madison and L. A. Steen, 27–44. Washington, DC: Mathematical Association of America. <http://www.maa.org/ql/cvc/cvc-027-044.pdf> (accessed May 25, 2011).
- Steen, L. A., ed. 2001. *Mathematics and democracy: The case for quantitative literacy*. Washington, DC: Woodrow Wilson National Fellowship Foundation, <http://www.maa.org/ql/mathanddemocracy.html> (accessed February 12, 2011).
- Taylor, J. 2001. Forcing fuel efficiency on consumers doesn't work. *Lincoln Journal-Star*, August 21, 2001. Reprinted in (Madison et al. 2009).
- Wiggins, G. 2003. Get Real! Assessing for Quantitative Literacy. In *Quantitative literacy: Why numeracy matters for schools and colleges*, ed. B. L. Madison and L. A. Steen, pp.121-143. Princeton, NJ: National Council on Education and the Disciplines.