Measuring Habits of Mind: Toward a Prompt-less Instrument for Assessing Quantitative Literacy

Stuart Boersma
Central Washington University, boersmas@cwu.edu
Dominic Klyve
Central Washington University, klyved@cwu.edu

Measuring Habits of Mind: Toward a Prompt-less Instrument for Assessing Quantitative Literacy

Abstract

In this study, we offer a new "prompt-less" instrument for measuring students' habits of mind in the field of quantitative literacy. The instrument consists of a series of questions about a newspaper article the students read. The questions do not explicitly solicit quantitative information; students' habit of mind is assessed by their use of quantitative reasoning even when it is not asked for. Students' answers were graded according to a modified version of the Quantitative Literacy Assessment Rubric (QLAR) published in this journal (vol. 4, issue 2). We applied the instrument and rubric to assess pre- and post-intervention habits of mind in opportunistic samples of two cohorts of students: the general (non-STEM) student body and (non-STEM) honors students at Central Washington University. The intervention was a QL course designed around a collection of newspaper articles to provide authentic context. The pre- and post-course assessment showed no statistically significant improvement in either group. We close with a discussion of practical aspects of using the rubric based on our experience of using it in this QL class.

Keywords

habit of mind, assessment, open-ended, reasoning, rubric

Creative Commons License

Cover Page Footnote

Stuart Boersma is a Professor of Mathematics at Central Washington University. He enjoys writing expository mathematics papers and received the 2005 Trevor Evans Award from the Mathematical Association of America.

Dominic Klyve is a an Assistant Professor of Mathematics at Central Washington University. In addition to his interest in effectively teaching quantitative literacy, he is interested in eighteenth century mathematics and computational number theory.

Introduction: The Need for Prompt-less Instruments

Descriptions of quantitative literacy (QL) often refer to a basic mathematical skill set uniquely combined with reasoning abilities, critical thinking abilities, and the habit of mind to purposefully engage with quantitative material. For example:

- *"the habit of mind to consider both the power and limitations of quantitative evidence in the evaluation, construction, and communication of arguments in public, professional, and personal life"* (Grawe 2011).

- The ability to *seek[s] out the world and use[s] quantitative skills to come to grips with its varied settings and concrete particularity"* (Robert Orrill in Steen 2001).

- *"a predisposition to look at the world through mathematical eyes, to see the benefits (and risks) of thinking quantitatively about commonplace issues, and to approach complex problems with confidence in the value of careful reasoning"* (Steen 2001).

- *"An aggregate of skills, knowledge, beliefs, dispositions, habits of mind, communication capabilities, and problem solving skills that people need in order to engage effectively in quantitative situations arising in life and work"* (Steen 2001).

- *"the power and habit of mind to search out quantitative information, critique it, reflect upon it, and apply it in their public, personal and professional lives"* from the vision of the National Numeracy Network as cited by Madison and Steen (2008).

The phrase "habit of mind" is frequently employed to capture the idea that quantitative literacy encompasses more than just the ability to respond intelligently to specific prompts. Rather, as the above quotations suggest, many posit that a quantitatively literate person will have a predisposition to employ a number of mathematical and critical thinking skills *on their own initiative* as opposed to simply responding to a series of prompts. As Grawe (2011) and Wiggins (2003) lament, this is precisely what makes *reasonable assessment* of such a definition of QL seemingly impossible. One must be able to design an instrument (presumably a series of prompts) in order to solicit the type of measurable response which will contain elements of a QL skill set. We use the phrase "reasonable assessment" as one could envision an experimental design in which subjects are carefully observed over a long time frame and interviewed in such a way as to identify when, where, how, and to what extent the subjects employed QL skills in their everyday lives. Such an extensive and intrusive study is not reasonable for most colleges, universities, or educators.

The "habit of mind" component of QL differentiates between *"can* my students ____" and *"do* my students regularly ____," where "____" could

include a multitude of quantitative reasoning activities. For the current study we propose to measure whether our students have the inclination to

- glean, identify and report quantitative information in direct support of a thesis statement;
- invoke quantitative reasoning to critique a statement or opinion;
- check numerical information presented in text with any accompanying graphics; and
- critically evaluate information presented graphically.

QL assessment is also problematic because of the propensity of instruments to focus too heavily on student responses, thereby diminishing the importance of both the reasoning that produces the response as well as the ability to communicate the response. This is certainly the case with multiple-choice assessments.

Currently available assessment instruments have not been found adequate for measuring the "habit of mind" component of QL. Kosko and Wilkins (2011) examined several open-ended assessments (from the Third International Mathematics and Science Study, TIMMS; the National Adult Literacy Survey, NALS; the International Adult Literacy Skills, IALS; and the Program for International Student assessment, PISA) and found that these tools had a tendency to assess answers as opposed to the reasoning required to arrive at those answers. Grawe (2011) claims that the Collegiate Learning Assessment (CLA) open-ended portion does not adequately measure a "habit of mind" to engage in QR because the prompts themselves contain enough hints or clues to bias students to consider quantitative reasoning approaches in their responses. The Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) institutional assessment initiative described by Grawe (2011) provides both a procedure and a useful rubric for programmatic QL assessment. Colleges looking to obtain institution-wide "snapshots" of the QL ability of their student body should consider the method described by Grawe (2011). However, for the individual instructor interested in assessing QL abilities of their students (before, during, and after a single course), a less time-intensive method would be desirable. The rubric we used for this study is based on the Quantitative Literacy Assessment Rubric (QLAR) which has been found to be a reliable scoring tool for student work (Boersma et al. 2011).

In order to measure the effectiveness of one component of our quantitative reasoning course at Central Washington University (CWU), we wished to administer a small-scale, open-ended QL assessment instrument in order to measure students' habit of mind to seek out and use quantitative information in an authentic contextual environment. We sought to design an instrument which could be administered in a single class period (50 minutes), be scored by a single instructor in a short amount of time, and lend itself to a pre- and post-intervention

assessment protocol. In addition to helping us assess our course, we were also very interested in identifying student learner differences between two populations of students at CWU: general non-STEM majors and those non-STEM majors enrolled in our honors program. In order to accomplish these goals we needed to design a prompt-less instrument and a companion rubric capable of measuring some aspects of the "habit of mind" component of QL. This paper will describe the instrument, the use of the companion rubric, a reflection on the actual implementation of the rubric in university classes, and the analysis of the data we collected in our classes.

A Framework for Quantitative Literacy

Our definition of quantitative literacy, and in particular our desire for students to demonstrate quantitatively literate habits of mind while reading media articles, is situated in the framework developed in Wilkins (2000). Using the goals of the Third International Mathematics and Science Study, together with earlier frameworks and research on scientific literacy, Wilkins proposed a framework that suggests a quantitatively literate person possesses:

- A functional knowledge of mathematical content.
- An ability to reason mathematically.
- A recognition of the societal impact and utility of mathematics.
- An understanding of the nature and historical development of mathematics.
- A positive disposition toward mathematics.

In the context of students' deciding on their own initiative to use quantitative reasoning while reading media articles, the first two components are implicitly necessary. The third necessitates, as Wilkins points out, that "being quantitatively literate … includes an awareness of the usefulness of mathematics and the ability use mathematics in everyday situations." This, of course, is precisely our goal.

Wilkins's fourth component seems at first glance to have very little to do with the reading of media articles. Part of the description of this quality, however, states that a student meeting this component would view mathematics "as a dynamic discipline characterized by inquiry and investigation instead of by procedures and memorization" (Wilkins 2000, p. 407). Our goal for the students is that they select appropriate tools of inquiry and investigation to approach each article – memorizing a single strategy is of little use.

Finally, a quantitative engagement with media articles requires that students "possess a willingness to engage in situations that require a functional level of quantitative reasoning" (Wilkins 2000, p. 408). This last idea, taken from Wilkins' description of his fifth component, serves neatly as a summary goal of

our class. Students must be *willing* to start to use their quantitative skills even when they are not being directly asked to do so.

Following Scheaffer's (2008) recommendation for research in quantitative reasoning, we aim to contribute an assessment instrument that will enable future researchers to continue to create more efficient, reliable, and valid assessment instruments of some habit of mind aspects of QL, to be tested in larger-scale studies among more general populations.

The Design of a Prompt-less Instrument

Newspaper articles can be an excellent source of contextually rich and quantitatively demanding material (e.g., Watson 2004; Dingman and Madison 2010; Madison and Dingman 2010; Boersma et al. 2011). Newspaper articles are situated in authentic contexts, written to be understood by a large percentage of our population, and are reasonably short. College-level mathematics courses can be designed around collections of carefully chosen articles and carefully designed problem sets (e.g., Dingman and Madison 2010; Madison and Dingman 2010; Boersma et al. 2011). Newspaper articles can also be used as a natural prompt-less QL assessment instrument! Ideally, one could choose an appropriate article, hand it out to students along with a blank sheet of paper and say "go!" Realistically, however, more direction will probably be needed in order to keep students from filling their papers with doodles, random thoughts, and the like. Thus, our final instrument is not truly a prompt-less instrument. However, the prompts were carefully created in order not to overtly lead students to provide the type of responses our rubric was designed to identify.

We began by choosing two newspaper articles. We decided that the articles should:

- Be roughly 500 words to allow students to read the articles in class and have enough time to complete the assessment.
- Contain content of interest to college students.
- Contain a variety of quantitative statements thereby requiring students to isolate those statements that are more central to the main theme of the article. Statements using relative quantities (percents, percentiles) and absolute quantities (specific counts) should also be present.
- State an argument(s) and use quantitative comparisons in support of the argument(s).
- Be accompanied by a graph which exhibits some discrepancies between the numerical information presented in the article.
- Be ripe for criticism – allowing for dialog on its strengths and weaknesses.

We chose (1) "Top students show little gain from 'No Child' efforts" by Liz Bowie as printed in *The Baltimore Sun* (June 18, 2008) and (2) "Tally high for Americans at Polls this year" by Fredreka Schouten as printed in *USA Today* (November 6, 2008). Although copyright restrictions prevent us from including the articles here, the text of the *Baltimore Sun* article, and the text and graph from the *USA Today* article, can be found with a simple web search.

When designing our prompts, we adhered to the following principles:

- Prompts are open-ended and not multiple choice.
- Prompts do not explicitly refer to any quantitative or mathematical calculation, requiring students to focus on those issues or ideas which they feel are most important.
- The instrument has only a short amount of space after each prompt, thus requiring students to distill their ideas to only a few sentences (some students did use the blank back of the instrument.)

Figure 1 shows the final form of the instrument.

Figure 1. The one-page instrument used for an accompanying newspaper article.

As mentioned above, it is clear that this instrument contains five prompts, and so it is not prompt-less. However, we assert that these five questions (primarily questions 3–5), together with our method of scoring student responses, still allow us to measure whether students have a "habit of mind" to (1) glean, identify and report quantitative information in direct support of a thesis statement; (2) invoke quantitative reasoning to critique a statement or opinion; (3) check numerical information presented in text with any accompanying graphics; and (4) critically evaluate information presented graphically. While the instrument itself does not specifically lead students to focus on quantitative information, we certainly acknowledge that, because the instrument was handed out in a quantitative reasoning course, students could easily infer that they should pay special attention to the numerical information in the article (although many did not). Additionally, since the post-intervention assessment was administered towards the end of a quantitative reasoning course which focused on the critical analysis of numerical information in newspaper articles, students should have been expected to key in on these characteristics. Indeed, this instrument provides a valuable and relevant assessment for this type of a course.

Methods

Because we were interested in measuring improvement in students' habit of mind to seek out, use, and critique quantitative information, we gave students both a pre- and post-intervention assessment. The initial assessment was given within the first three days of class. Students were told it would not be graded for correctness, but successful completion counted as a homework assignment, thus providing incentive to take the assignment seriously. Individual student papers were not handed back, but some class time was devoted to short discussions based on the article and assessment assignment. The rest of the course was spent critically analyzing newspaper articles with a quantitative reasoning lens (the intervention). The final (post-intervention) assessment was administered during the last week of class. As with the pre-intervention assessment, students were told that the papers would not be graded on correctness, but successful completion would count towards their homework score.

We used this methodology in three classes comprised of two "cohorts" of students. One class was taught in Fall 2009 (by SB), consisted of 23 non-STEM majors from the general population of students at Central Washington University, and used the textbook by Madison et al. (2009). Because these 23 students came from our university's general population, they will be identified as Cohort G for this study. Students self-selected this particular course.

Two other classes were taught by the authors (SB and DK) in Fall 2011 to non-STEM majors at CWU who were enrolled in the university's Honors College.

These classes also used the textbook by Madison et al. (2009) and consisted of 20 students each. Because these 40 students came from the Honors College, they will be identified as Cohort H. Again, these students self-selected the specific section of the course based on their individual schedules. There was no attempt at randomization.

The Use of the Prompt-less Instrument

Boersma et al. (2011) describe a Quantitative Literacy Assessment Rubric (QLAR) that can be used to assess individual student work in quantitative literacy. The QLAR assesses student performance in six core competencies: Interpretation, Representation, Calculation, Analysis/Synthesis, Assumptions, and Communication. These six competencies first appeared in the VALUE rubrics (Rhodes 2010) from the Association of American Colleges and Universities (AAC&U). We used the QLAR to help us score student responses. Specifically, we focused on the core competencies of Interpretation and Analysis/Synthesis and made slight adjustments to the rubric descriptors in order to tailor the QLAR to this specific project.

Language and Reading

The first question – "Did you understand the Article?" – was used to help identify those students who may have difficulty with the language encountered in the reading. If a student self-identifies as having difficulty with the language, then the assessment data gathered may not be informative. No student self-identified as having such difficulties.

The second question – "What was the main point(s) of the article?" – measures a student's ability to identify the main point of a lengthy article. Because these are newspaper articles, the main point(s) are often clearly identified by reading the headline, sub-headline, and first sentence of the article (this was not, however, recognized by all students).

Habit of Mind to Seek out and Report Quantitative Information

The third question – "What facts did the author use to support the main point(s)" – is the first to address the habit of mind component of QL. Scoring of this item focused on whether or not students could glean, identify, and report all the relevant quantitative information used in direct support of the main point (each article contained much more quantitative information than needed to be reported). This is essentially the core competency "Interpretation" of QLAR. To score high on this item, students must (1) have the habit of mind to seek out quantitative

information in the article, (2) identify relevant and specific information, and (3) communicate these facts in one to three sentences. For this item, the following rubric was used to score student responses (sample student solution in italics):

Score 0: No quantitative information given or alluded to. *"A study by the Brookings Institution."*

Score 1: Some relevant quantitative information is identified (or alluded to), but none is correct (or specific enough to be judged correct or incorrect). *"The average increase in NAEP test scores for lower and top students, teacher and public responses, and quotes from school staff."*

Score 2: Some relevant and correct information is identified, but not all. *"The lowest performing gained 22 points in 7 years while the highest gained 9 points."*

Score 3: All relevant quantitative information is correctly identified. *"The nationwide fourth-grade reading scores for the poorest-performing students have risen 16 points since 2000 compared with only 3 points for the top students. A national teacher survey showed that 60% of teachers said that the struggling students were the top priority in their school."*

Clearly this question prompts students for supporting facts. However, it does not specifically prompt them for facts that are both (1) relevant and (2) quantitative in nature. While the article contains non-quantitative facts that support the main points, our scoring methodology does not reward these responses. This is not because they are inappropriate responses for the given prompt, but because they do not reflect the habit of mind to seek out and report quantitative information. We are measuring the presence of a very specific habit of mind, not grading student responses.

The fourth question – "Were there any particular strengths or weaknesses in how these facts were reported?" – measures whether or not students focus on quantitative strengths or weaknesses as opposed to presentation (use of bullets, photographs, author writing style, etc.). For this item students needed to (1) draw a conclusion (are there any strengths or weaknesses present in the article? and (2) base their conclusion with correct and complete quantitative analysis. This is essentially the core competency "Analysis/Synthesis" of QLAR. Scoring of this item proceeded as follows:

Score 0: No strength or weakness identified or, if identified, not supported with quantitative reasoning. *All the numbers got kind of confusing.*

Score 1: A strength or weakness is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to

be able to judge correctness). *They could have compared voter turnout to the '04 election better by absolute population.*

Score 2: A strength or weakness is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g., it contains unsubstantiated claims). *The 62.5% could be misinterpreted as the estimate of votes for Obama.*

Score 3: A strength or weakness is identified and supported with correct and complete quantitative reasoning. *A strong weakness was that this data was calculated before official results, so these may not be the true numbers. A strength is that they presented the data in 2 different ways: in solid numbers and in percentages.*

Again, while we are prompting students to focus on the strengths and/or weaknesses present in the article, we are not prompting them to support their conclusions with quantitative analysis. Our method of scoring was chosen to measure students' habit of mind to engage in quantitative reasoning on their own initiative – essentially our definition of "habit of mind".

The fifth question – "Does the graph help interpret the numerical information found in the text? Explain your thoughts." – was used to measure (1) a student's habit of mind to check the numerical information in the text with the numerical information being presented graphically and (2) students' habit of mind to critically evaluate graphical information within an authentic context. Student responses were scored twice. During the first pass, we used the following scoring guide:

Score 0: No indication that the numbers in the article were checked against their representation in the graph. *Somewhat. The first graph supports what is mentioned in the article and the second graph helps to prove the same point.*

Score 1: Claims, with no justification, or incorrect justification, that the graph does or does not accurately present the numerical information in the article. *Yes because it accurately compares the two levels [of] progression over 11 years.*

Score 2: Claims, with justification, that the graph does or does not accurately present the numerical information found in the text. *Yes because it shows how the 90% students didn't really improve over 4 years and the 10th percentile students did.*

Score 3: Correctly points out a specific discrepancy between the graphical presentation and the quantitative information found in the text. *Not really. It seems like pretty weak support to me the more I look at it. And the 16*

points & 3 points don't really have a place on this graph. The points sound like a whole new graph.

While this fifth question is the most leading on the instrument, we believe that it is open-ended enough to allow us to measure the students' habit of mind to check specifically to see whether the numerical information presented in the text differs substantially from the accompanying graphic. Both articles were carefully chosen to allow for astute students to be able to identify such discrepancies *if they had the predisposition to look for them.*

In order to measure the students' habit of mind to critically evaluate graphical information within an authentic context, we were guided by the "Analysis/ Synthesis" competency of QLAR. We were looking to see if students could draw a conclusion (regarding the usefulness of the graph) and support their conclusion with quantitative analysis. During the second pass, we used the following scoring guide:

Score 0: No strength or weakness of the graph identified or, if identified, not supported with quantitative reasoning. *Yes, the graph does help, but the quantitative information was presented clearly enough to understand without the graph. It is a nice visual aid, however.*

Score 1: A strength or weakness of the graph is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness). *The x-axis is really weird. They should have just used a bar graph or pie chart showing 1960, 1968, and 2008 voter turn out. To have a more compelling chart.*

Score 2: A strength or weakness of the graph is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g. it contains unsubstantiated claims). *It nicely shows that the number of those that have voted has indeed increased. But having the years skip at the bottom is somewhat annoying, I'd rather they keep it consistent.*

Score 3: A strength or weakness of the graph is identified and supported with correct and complete quantitative reasoning. *It certainly does seem to support the claim that voting turnout in 2008 was 62.5% and that in 1968, 63.8% of people voted. The 51.7% seemed a little random till I realized it's probably the lowest point between these two years. They are also nice enough to write at the bottom that this is an "unofficial estimate".*

While students were prompted to focus their discussion on the graphic which accompanies the article, our scoring methodology identified students who took the initiative (unprompted) to comment on any particular strengths or weaknesses they noticed in the graph.

Analysis

In this section, we wish to do two things. The first is to answer our content course-assessment questions, to determine whether our course did improve students' habit of mind to use quantitative reasoning, and to determine whether cohort H responded differently to instruction than cohort G. Our second goal is to describe and reflect on the use of this instrument in actual classroom practice.

We also note here a significant limitation to our analysis. We have no way of knowing whether it is equally challenging for students to quantitatively asses the two different newspaper articles. If finding discrepancies between the data and the graph, for example, is much trickier for one of our two articles, then the pre- and post-intervention data are not measuring what we are after. Because both cohorts used precisely the same two articles for their assessments, however, there is no difficulty in comparing the improvement between the two cohorts.

To repeat, the assessment was given twice during each of our QL classes: once during the first three days of class, and once during the last week of class. The pre- and post-coursework scores for each of our six assessment questions are given for our two cohorts in Table 1.

Clearly the honors students in Cohort H had considerably higher scores on the pre-course test for most questions, reflecting their stronger background. The only exception was question 3 ("What facts did the author use to support the main points?"). In the first article, it seems that the quantitative supporting data were presented in such a way that stronger students had no significant advantage in identifying it.

Table 1
Average scores on each question for the given cohorts

Question	Cohort G		Cohort H	
	Pre-	Post-	Pre-	Post-
3	1.52	1.21	1.53	1.95
4	0.33	0.16	1.29	1.08
5.1	0.81	0.84	1.42	1.49
5.2	0.05	0.53	0.47	0.3

Of more interest than the raw scores is the improvement the students made. Although our instrument seemed to be successful in measuring students' habit of mind, our course was less successful in actually changing these habits. Table 2 gives the average improvement made by students in each cohort for each question.

Table 2
Improvement between pre- and post-assessments for each question

Question	Cohort G	Cohort H	Overall
1	0	0	0
2	-0.01	-0.04	-0.03
3	-0.31	0.42	0.17
4	-0.17	-0.21	-0.18
5.1	0.03	0.07	0.07
5.2	0.48	-0.17	0.06

Here we see the disappointing result that, within the limits of our test, we have no evidence that students' habits of mind were significantly improved by

this course. None of the differences in the "Overall" column are statistically significant.

Only two of the improvements shown in Table 2 are statistically significant at the $p = 0.05$ level. Matched pairs t-tests give that question 3 for Cohort H and question 5.2 for Cohort G were both significant. Neither of these, however, remains significant after adjusting significance levels for multiple comparisons via, say, Bonferroni or Tukey methods. In terms of average score, therefore, no group showed significant change on any question. Similarly, there seems to be no statistically significant difference in improvement between Cohort G and Cohort H for any of the questions.

We are left with the difficulty that any differences in pre- and post-course scores on this assessment are possibly confounded by the fact that the students read two different articles for the two assessments. Lacking any objective measure that one article is inherently "easier" than the other to process, we are treating them as equal in the analysis above. For the purpose of this paper, the possible improvement or lack of improvement with our course is a side issue. We still find that we can measure students' habits of mind and quantitative reasoning abilities with our instrument, although it is clear that one would need to find a way to more precisely compare the pre- and post-intervention assessments for a better understanding of student learning. Two obvious possibilities for future work are randomized design, in which half of the students receive the articles in the opposite order, and an external measure to calibrate the level of difficulty of each.

Finally, we wished to assess the import of question 2: "What was the main point(s) of the article?" This question did not measure any type of QR competency, but we might still expect that students' ability to answer this question may have some effect on their answers to other questions. In fact, their answer had no significant effect on their answers to question 3, or question 5 (measured in either way), but there was a difference in students' score on question 4, "Were there any particular strengths or weaknesses in how these facts were reported?" (Table 3)

Table 3

Mean score on Question 4 for students who incorrectly (0) or correctly (1) identified the main point of the article in Question 2

Question 2	Question 4	N
0	0.60	50
1	1.06	65

Of all the questions on the instrument, question 4 may be the most open-ended. Where question 3 asks for facts (possibly hinting at the desire for quantitative figures), and question 5 asks about the graph (a quantitative tool), question 4 simply asks for overall strengths and weaknesses. It is interesting that students who grasp the main purpose of the article (a non-quantitative skill) are also the most proficient at choosing to apply quantitative reasoning when none is asked for.

Application of the Rubric

The primary purpose of this paper is to introduce an instrument designed to test students' habit of mind in using and applying quantitative reasoning. We would like, therefore, to reflect on the use and scoring of this instrument.

Each of the questions on every student's assessment was scored by each author separately. We removed the students' names from the top of the assessments and replaced them with unique subject numbers. Thus, when we applied the rubric to the students' work, neither author knew which student he was scoring, or indeed in which section the student had enrolled.

In practice, applying the rubric was perhaps a bit slower than traditional grading, although the difference was slight. We suspect with continued practice this small time penalty would disappear.

After agreeing on the rubric, we scored about five papers together in an attempt to build consensus for how precisely to apply it. After that, scoring was done independently. When we each completed our work, we met to discuss any discrepancies we found in our scoring. We found that we were able to reach a consensus on the best score fairly easily in these cases.

For most of the problems, we found we disagreed in 20–25% of the cases. A notable exception occurred in scoring the second pass at question 5, measuring the "Analysis/Synthesis" competency. For this problem, we found that our scores had initially disagreed in more than fifty percent of cases. These differences, and indeed most of the differences in the other scores, were due to different interpretations of imprecise parts of the rubric. We used our experience in sorting through these differences to refine the rubric, and believe it can now be used with high reliability.

Conclusions

The many descriptions and definitions of quantitative reasoning which appear at the beginning of this article are reminders of two of the facets that make quantitative reasoning such a challenging and rewarding field to work in, namely that QL is difficult to teach, and it is difficult to assess. We believe that the instrument described in this paper will be of significant help in meeting the second of these challenges. We plan to continue to use this instrument as we calibrate our own teaching and work toward more effective methods for teaching new habits of mind to our students.

Acknowledgments

References

Boersma, S., C. Diefenderfer, S. Dingman, and B. 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

Bowie, L. 2008, Top students show little gain from 'No Child' efforts, *The Baltimore Sun*, June 18, 2008.

Dingman, S. and B. 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

Grawe, N. 2011. Beyond math skills: Measuring quantitative reasoning in context. *New Directions for Institutional Research* 149: 41–52. http://dx.doi.org/10.1002/ir.379

Kosko, K. and J.L.M. Wilkins. 2011. Communicating quantitative literacy: An examination of open-ended assessment items in TIMSS, NALS, IALS, and PISA. *Numeracy* 4 (2), Article 3. http://dx.doi.org/10.5038/1936-4660.4.2.3

Madison, B. and S. Dingman. 2010. Quantitative reasoning in the contemporary world, 2: Focus questions for the numeracy community. *Numeracy* 3 (2). http://dx.doi.org/10.5038/1936-4660.3.2.5

Madison, B. and L. Steen. 2008. Evolution of numeracy and the National Numeracy Network. *Numeracy*, 1 (1), Article 2. http://dx.doi.org/10.5038/1936-4660.1.1.2

Madison, B., S. Boersma, C. Diefenderfer, and S. Dingman. 2009. *Case Studies for Quantitative Reasoning: A Casebook of Media Articles*, 2nd ed. New York: Pearson Custom Publishing.

Rhodes, T.L., ed. 2010. Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics. Washington D.C. Association of American Colleges and Universities.

Scheaffer, R.L. 2008. Scientifically based research in quantitative literacy: Guidelines for building a knowledge base. *Numeracy* 1 (1), Article 3. http://dx.doi.org/10.5038/1936-4660.1.1.3

Schouten, F. 2008. Tally high for Americans at polls this year, *USA Today,* November 6, 2008.

Steen, L.A. ed. 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: The National Council on Education and the Disciplines.

———. 2004. *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education*. Washington D.C.: The Mathematical Association of America.

Watson, J. 2004. Quantitative literacy in the media: An arena for problem solving. *The Australian Mathematics Teacher* 60, Part 1: 34–40.

Wiggins, G. 2003. Get Real!: Assessing for quantitative literacy. In B.L. Madison and L.A. Steen (eds), *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges*. Princeton, NJ: National Council on Education and the Disciplines.

Wilkins, J.L. 2000. Preparing for the 21st Century: The status of quantitative literacy in the United States. *School Science and Mathematics*, 100 (8): 405–418. http://dx.doi.org/10.1111/j.1949-8594.2000.tb17329.x