Keynote Presentation
Reflections on Quantitative Reasoning:  
An Assessment Perspective  

Richard J. Shavelson*  
*Stanford University  

If we seek to enhance the quantitative reasoning of the American public, not only do we need to be able to say what quantitative reasoning is, we also need to know how to teach for it and how to measure progress toward the QR goal. My focus is on assessment, but inevitably I also need to address definitional questions. Moreover, teaching and assessing go hand-in-hand. So what is said about assessment has application to teaching (and vice versa).

From an assessment perspective the first question that arises about QR is: “What is the construct to be measured?” Or, “What is the ‘theory’ or ‘model’ of QR from which an assessment emanates?” This question conjures up over 100 years of study of quantitative reasoning. Psychometric, cognitive, and situated theories all have something to say about the question. An early task, therefore, is to set forth a simple assessment framework that provides the structure for the paper—what is to be measured, with what tasks, and with what inferences? Once the framework is sketched, I will introduce definitional questions and set forth a particular definition that constrains what we might assess and how we might assess it.

*Richard J. Shavelson is the Margaret Jack Professor of Education, Professor of Psychology, Senior Fellow in the Stanford Institute for the Environment, and former I. James Quillen Dean of the School of Education at Stanford University. He served as president of the American Educational Research Association; is a fellow of the American Association for the Advancement of Science, the American Psychological Association, and the American Psychological Society; and is a Humboldt Fellow. He is completing a book entitled (tentatively): The Quest to Assess Learning and Hold Higher Education Accountable. His other books include Statistical Reasoning for the Behavioral Sciences, Generalizability Theory: A Primer (with Noreen Webb), and Scientific Research in Education (edited with Lisa Towne). E-mail: richs@stanford.edu
With a working definition of QR in hand, the assessment question turns to what kinds of tasks might be used to measure the construct. Two approaches to measuring QR—one with roots in behaviorism and the other from cognitive and situative perspectives—illustrate different underlying conceptions of learning, knowing, and performing. They lead to different answers to the question of what tasks should be used to measure QR. As we will see, certain kinds of tasks are likely to elicit quantitative reasoning in a manner consistent with our definitional view, while other kinds of tasks are less likely to do so. One approach that I like—akin to cognitive/situative orientations—is reflected in the new Collegiate Learning Assessment (CLA). Consequently, I describe it in some detail to illustrate one important direction for assessment of QR that may fit what the field is looking for.

With the assessment built, the final question that arises is, “How justifiable is the inference from test information to students’ or teachers’ level of quantitative ability?” Due to space limitations, I only note but do not discuss the need to amass empirical evidence and the kinds of evidence that support a QR interpretation. In concluding, I examine from a variety of perspectives the potential for teacher preparation and enhancement to improve QR, thereby raising questions as to whether teacher education in QR is the most effective approach to deal with this 21st century challenge.

Approaching quantitative reasoning through assessment

There is a growing consensus that to function effectively in the 21st Century, Americans need to be “quantitatively literate,” that is, be able to think and reason quantitatively when the situation so demands. And by implication, Americans certainly need to be able to quantitatively reason better than they can today. This view is clearly expressed in the invitation letter for this conference: “The goal of the conference is to explore educational solutions to the increasing quantitative reasoning demands on US residents.”

If we seek to enhance the quantitative reasoning of the American public, we need not only to be able to say what quantitative reasoning is, we also need to know how to teach for it and how to measure progress toward the QR goal. This paper focuses on assessment of quantitative reasoning. But assessment, strangely enough, is a good road into the topic because it forces us to be clear about what we mean by QR, what kinds of tasks or activities would elicit QR (both in the classroom and on the assessment), and what kinds of evidence is needed to convince ourselves we are measuring the “right stuff.”

Perhaps surprisingly, the way assessment developers approach the development of measures of any construct like QR is instructive not just for
assessments, but also for teaching and learning. Development work is guided by an assessment triangle (NRC, 2001) shown in Figure 1. The construct vertex represents the “thing” or “concept” or “construct” we want to measure. In our case, the construct is quantitative reasoning. That is, we want to infer the level of quantitative reasoning displayed by an individual or group of individuals based on our assessment. To do so means we have to begin by defining what we mean by QR. Such a definition is not set in stone. Rather, as we gain experience teaching QR and empirical evidence about the adequacy of our assessment of QR, we may very well modify the definition. But such a working definition is a starting point.

The observation vertex of the triangle represents the kinds of activities we believe would permit an individual or group to display QR. The definition of the construct helps define a universe of possible activities—tasks and how they might be responded to—that we might use to assess QR… or to teach QR! The definition also rules out some activities that we would not consider as counting as eliciting the kind of QR we have in mind. Typically we do a kind of task analysis to insure, at least logically, that the tasks/ responses that form the activities on the assessment are drawn from the universe of QR activities that we intend to draw inferences to, based on assessment scores.

Finally, the interpretation vertex focuses on the inferences we make from a sample of activities to the universe of activities that we want to know about a person in order to capture his or her QR. By interpretation is meant the basis for scoring performance and the chain of reasoning—logical, cognitive, and statistical—that links the scores on the assessment to the construct of interest, QR. Indeed, we do not know what an assessment measures unless we know what the tasks are, how people are asked to respond to those tasks, and how those responses are scored. And even then we need logical, cognitive, and statistical evidence that supports our interpretation that we are really measuring the QR we set out to measure. Given limited time, this vertex will not be discussed herein further.

I hope by now I have convinced you that the assessment process demands a great deal of reasoning, especially quantitative reasoning!
Definitions of quantitative reasoning

Historically, within psychology and education, there have been three approaches to defining QR: psychometric (behavioral roots), cognitive (mental process roots) and situative (social-contextual roots). Each sheds light on what we might or might not mean by QR. And, as we will see, current definitions of QR in the mathematics QR community overlap some combination of the cognitive and situative here.

Psychometric Approach

The psychometric approach begins with a “mini-theory” of what QR might be and then builds tests to match that theory. It then tests the theory empirically, looking for patterns of correlations among test scores such that tests measuring QR should correlate higher with each other than with tests of, say, verbal or spatial reasoning. This tradition has been ongoing for more than 100 years. There seems to be consensus that there is strong evidence for a “QR factor” in the sense that people’s performance on QR tests can be distinguished clearly from their performance on other tests. QR “… requires reasoning based on mathematical properties and relations. The reasoning processes may be either inductive or deductive, or some combination of them” (Carroll, 1993, p. 239). QR tests have titles such as “Arithmetic, Necessary Arithmetical Reasoning and Mathematical Aptitude.” Carroll goes on:

Typically these tests present a variety of mathematical reasoning problems such as word problems (solving verbally stated mathematical problems), number series, and problems requiring selection of appropriate arithmetical operations. Generally, the amount of actual numerical computation required is small. [S]cores are expected to depend mainly on the level of difficulty in the problems that can be performed.

To put QR in context, a figure generated by Snow and Lohman (1989, p. 318, Figure 3.13) and adapted by Gustafsson and Undheim, (1996, p. 201, Figure 8-5) is helpful (Figure 2). This “dartboard” representation of human cognitive abilities shows the bull’s eye to be general mental ability. Radiating out from the center are verbal, spatial, and quantitative reasoning. Let us focus on the QR piece of the board (or slice of pie). As we move away from the bull’s eye toward the edge of the board, the tests of QR become increasingly like those tasks that might be taught in school and, consequently, most influenced by education. This said, those tests closer to the bull’s eye seem to best reflect what psychometricians think of as QR.
To get a better feel for some of these tests, questions from two QR tests are presented in Figures 3a, 3b, and 3c. The test question in Figure 3a, taken

Example II. Chairs priced at $40 each are being sold in lots of 4 at 85% of the original price. How much would 4 chairs cost?

1 - divide and add 3 - subtract and divide
2 - multiply and multiply 4 - multiply and divide

One way to solve the problem would be to multiply $40 by .85 and then multiply this product by 4; therefore you should have put an x through the number 2. (Although some problems may be solved in more than one way, as with Example II, only the operations of one of these ways will be given among the options).

When 2 operations are given, they are always given in the order in which they should be performed.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.
Calculation vs. Context

from the Necessary Arithmetic Operations Test (Ekstrom, French, Harman, & Dermen, 1976), falls close to the bull’s eye in Figure 2 while the test question in Figure 3b, taken from the Arithmetic Aptitude Test (Ekstrom, French, Harman, & Dermen, 1976), falls just beyond. For completeness, Figure 3c shows the Addition Test (Ekstrom, French, Harman, & Dermen, 1976) which falls at the periphery of the dart board and is closely tied to education and practice.

The questions in Figure 3a and b, then, appear to be consistent with Carroll’s claim that the tests do not place a high demand on computation but rather focus on reasoning with numbers, operations, and patterns. The last question focuses on numerical speed and accuracy, not what is meant by QR in the psychometric view of cognitive abilities.

You are to write your answers in the boxes below the problems. Several practice problems are given below with the first one correctly worked. Practice for speed on the others. This practice may help your score.

Example: How many candy mints can you buy for 50 cents at the rate of 2 for 5 cents?

1- 1 2- 20 3- 25 4- 100 5- 125

The correct answer to this problem is 20. Therefore, you should have marked an x through the number 2 to indicate the correct answer.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

Cognitive Approach

Psychometricians have focused on observed performance or behavior in response to a set of similar test questions seeking to understand the structure of
abilities underlying consistency in responses to these questions. In contrast, cognitive scientists assume that performance on cognitive tests can be divided into component processes and ask about the cognitive operations that underlie the observed performance on psychometric and other tasks. Or, as John Anderson put it, “Most of the research in psychometrics has focused only on whether a person gets a question right or not. In contrast, information-processing [cognitive] analyses try to examine the steps by which a person decides on an answer to such a question and the time necessary to perform each step” (2005, p. 447). Cognitive scientists’ goal is to extract sets of elementary processes that underlie a wide range of cognitive functioning and thereby describe elemental cognitive processes.

With respect to QR, cognitive scientists (actually seldom) ask, “What kinds of reasoning processes or steps are brought to bear in responding to QR type tasks? In what order? And for how long?” Unfortunately, QR has not been a focus of much cognitive research (although deductive and inductive reasoning have). One possible example (Figure 4) is cryptarithmetic (Newell & Simon, 1978, p. 143):

Although not the best example of QR, cryptarithmetic does provide a sense of the cognitive scientists’ approach in analyzing a potential QR task. They begin by setting forth the task environment—the affordances and constraints of the problem—for a single problem. They then analyze, logically and mathematically, the possible solution paths for the problem. The next step in the analysis is to move from this intensive analysis of a single task to an extensive analysis which generalizes the rules for solving one problem to other problems that fall in the same domain. They then observe human performance on the task, asking problem solvers to “think aloud” to capture the reasoning processes underlying task completion. In this way, they map the “problem space” that the problem solver has constructed and the step by step processes used in problem solving. They may then test their conclusion by building and testing computer models of problem solving.

This approach, then, has less to say about constructing assessment activities than it does about how to determine whether the assessment activities tap the

---

**Figure 4. Cryptarithmetic**

```
DONALD   D = 5
+ GERALD
ROBERT
```

Here each letter represents a digit (0, 1,…,9) and you know D = 5; no other letter equals 5. What digits should be assigned to the letters such that, when the letters are replaced by their corresponding digits, the sum above is satisfied?
kind of thinking—QR—of interest. Cognitive methods, such as the think-aloud technique, provide important means for examining proposed interpretations of assessments purporting to measure QR.

**Situated Approach**

While psychometricians ask, figuratively, “How fast will the car go?”, and cognitive scientists ask “How does the engine make the car go fast?”, situativists ask “How is the car used in a particular culture?” Situativists ask about person-in-situation. They view performance as influenced in part by what the individual brings to a situation and in part by the physical and social situation—its affordances and constraints—in which that performance becomes meaningful. In their pursuit of understanding human abilities, including QR, they also want to know how a particular culture affects the development and use of these abilities.

Indeed, situativists would probably frame the question of understanding QR a bit differently than has been done here. They would begin by not assuming that QR resides solely within the person but would view QR within a community of practice—e.g., those individuals engaged in culturally relevant activities in which reasoning quantitatively is demanded and the various resources of the community would be brought to bear on those activities. They would view a person accomplished in QR as having the capacity to engage others in working together to think critically, reason analytically and to solve a problem, for example. Cognitive abilities, from this perspective, reside in a community of practice.

To pursue the situative perspective further is a task for another time, as the capacity to assess performance poses a very real challenge for this perspective. And issues of credibility arise when those outside the situative community of practice are asked to buy into the way they assess performance.

That said it is possible to conceive of tasks that fit to some degree with this perspective. For example, the use of case studies in business which among other things demand QR, as Corrine Taylor (2007) points out, seems consistent with the situative perspective. QR is embedded in the larger set of real-world constraints and affordances and the problem solution depends upon them. Moreover, Bernie Madison’s (2006) characterization of QR in contrast to mathematics resonates with this perspective (Figure 5). QR, from his perspective, is carried out in real-life, authentic situations; its application is in the particular situation, one dependent upon context including socio-politics. The problems are ill defined, estimation is crucial, and an interdisciplinary approach is often needed.
Perhaps the following question, from Friedman’s statistics book, also falls within the situative perspective and what it means to reason quantitatively:

One of the drugs in the Coronary Drug Project was nicotinic acid. Suppose the results on nicotinic acid were as reproduced below. Something looks wrong. What, and why?

<table>
<thead>
<tr>
<th>Nicotinic Acid</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td><strong>Number</strong></td>
</tr>
<tr>
<td>Adherers</td>
<td>558</td>
</tr>
<tr>
<td>Non-adherers</td>
<td>487</td>
</tr>
<tr>
<td>Total</td>
<td>1045</td>
</tr>
</tbody>
</table>

*Answer:* About half of those getting Nicotinic Acid adhered to their treatment regimen whereas two-thirds of those getting the Placebo adhered to their regimen. This suggests something went wrong. For instance, the Nicotinic acid may have had unpleasant side effects or the Placebo tasted better. In short, it may not have been a true placebo.

Note that the exercise does not involve formulas (other than noticing the large difference in adherence rates).

These examples of situated QR seem also to fit with the Mathematical Association of America’s notion of QR; all students who receive a bachelor’s degree should be able to:
• Interpret mathematical models such as formulas, graphs, table, and schematics, and draw inferences from them.
• Represent mathematical information symbolically, visually, numerically, and verbally.
• Use arithmetical, algebraic, geometric, and statistical methods to solve problems.
• Estimate and check answers to mathematical problems in order to determine reasonableness, identify alternatives, and select optimal results.
• Recognize that mathematical and statistical methods have limits.

The definition that seems most productive from a present day notion of QR is that of the situativists, perhaps augmented by a cognitive analysis. With this in mind, approaches to measuring QR provide striking contrast.

A possible approach to measuring quantitative reasoning

The tasks used to elicit quantitative reasoning and how that reasoning is expressed in overt performance derives from the construct definition. The psychometric perspective’s definition of QR is helpful in thinking about the kinds of activities that might be included on a QR assessment—activities that require reasoning based on mathematical properties and relations. However, this perspective’s translation of the definition into assessment activities is constrained by the behaviorist notion that a complex task can be divided into component parts and then put back together again. Consequently, the questions found on psychometric tests are pretty much context free and posed in the form of multiple-choice test questions, as we saw in Figures 3a–c. This claim is reinforced by the GRE QR section (ETS, 2002). Carroll (1993) considers the GRE Quantitative scale to be prototypical of QR.

The kind of question in Figure 5, typical of the approach to measuring QR in the U.S., appears on many of the 30 or so QR websites I looked at having been provided a mere 1,800,000 to consider by Google. These questions appear

The average (arithmetic mean) of $x$ and $y$ is 20. If $z = 5$, what is the average of $x$, $y$, and $z$?

A. 8 ½  B. 10  C. 12 ½  D. 15  E. 17 ½

Answer: Since the average of $x$ and $y$ is 20, $(x + y)/2 = 20$ so $x + y = 40$. Thus $x + y + z = 40 + 5 = 45$, and therefore $(x + y + z)/3 = 45/3 = 15$.

Figure 5. Problem-solving question from the GRE released questions.
to draw on some aspects of QR. However, such questions, by their content and format (multiple-choice), seem context free in nature with one correct answer actually provided among a set of alternatives. Such an approach does not appear to be what many faculty and the MAA expect to see on a test of QR.

Rather, the situated approach seems to capture current thinking about QR. That is, QR is evidenced when confronted with a well contextualized, messy, open-ended, “real-world” task that demands analysis, critical thinking, problem solving and the capacity to communicate a solution, decision, or course of action clearly in writing. For example, two pieces of information provided in Figure 6 are part of an “in-basket” of information given to the problem-solver. The task smacks of the “real world” with substantial contextualization. The evidence points to a possible correlation between growth in sales of the SwiftAir aircraft and an increase in accidents—was the increase proportional? There are a number of solution paths and more than one solution to the problem could be justified.

The Collegiate Learning Assessment (CLA) provides one possible example of an assessment that fits a situated notion of QR. It poses complex tasks, provides a variety of information (e.g., data, graphs, research review article, newspaper article, op ed piece) and asks students to review the material, determine what material is relevant and what irrelevant, and arrive at a problem solution, decision, or course of action that is justified based on the evidence in hand. There is no single correct answer but a variety of possible answers that vary in their credibility and evidentiary base.

Given all of the information in the document library, what do you think are the three most likely causes of the accidents described in Document 3? Justify your answer with information from the document library.

Source: Collegiate Learning Assessment

Figure 6. Possible situated QR problem—two pieces of information regarding SwiftAir sales and accidents as part of an in-basket of information.
Because the CLA comes close to the situated definition of QR and what might be sought in an assessment of QR, the next section describes the CLA in some detail. In passing, note that the kinds of tasks used on the CLA would make excellent teaching activities. If faculty used these activities as part of their teaching, students might be more likely than at present to improve their reasoning. One caveat is in order. The CLA contains a number of performance tasks that demand among other things QR. But it is not a test of QR. The CLA is presented here as an example of an assessment that might very well be adapted to focus on QR. That said, philosophically it fits well with situated QR notions in that QR is more than quantitative reasoning; it involved an entire complex of reasoning and so do the CLA tasks.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Attributes</th>
</tr>
</thead>
</table>
| Open-ended Tasks        | • Tap critical thinking, analytic reasoning, problem solving and written communication  
                          | • Realistic work samples                                                  
                          | • Engaging task as suggested by alluring titles such as “brain boost,” “catfish,” “lakes to rivers”) | • Applicable to different academic majors |
| Computer Technology     | • Interactive internet platform                                           
                          | • Paperless administration                                                |
                          | • Natural language processing software for scoring students written communication |
                          | • Online rater scoring and calibration of performance tasks              |
                          | • Report institution’s (and subdivision’s) performance (and individual student performance confidentially to student) |
| Focus                   | • Institution or school/department/ program within institutions           |
                          | • Not on individual student performance (although their performance is reported to them confidentially) |
| Sampling                | • Samples students so that not all students perform all tasks            |
                          | • Samples tasks for random subsets of students                            |
                          | • Creates scores at institution or subdivision/program level as desired (depending on sample sizes) |
| Reporting               | • Controls for students’ ability so that “similarly situated” benchmark campuses can be compared |
                          | • Provides value added estimates—from freshman to senior year or with measures on a sample of freshmen and seniors |
                          | • Provides percentiles                                                   |
                          | • Provides benchmark institutions                                         |

Figure 7. Characteristics of the Collegiate Learning Assessment
The roots of the CLA can be traced to progressive notions of learning, focusing on critical thinking, analytic reasoning, problem solving, and written communication (Figure 7; Shavelson, 2007a,b). These capabilities are tapped in realistic “work-sample” tasks drawn from education, work, and everyday issues. They are accessible to students from the wide diversity of majors and general education programs. The capacity to provide these rich tasks without overburdening students is afforded by recent developments in information technology. The assessment is delivered on an interactive internet platform that produces a paperless, electronic administration and online report of results. Written communication tasks are scored using natural language processing software and performance tasks are currently scored by online human raters whose scoring is monitored and calibrated. Within the next year, the performance tasks will be scored as well by computer software.

The assessment is divided into three parts—analytic writing, performance tasks, and biographical information—the first two of which are relevant to present discussion. Two types of writing tasks are administered. The first, make an argument, invites students to present an argument for or against a particular position. For example, the prompt might be: “In our time, specialists of all kinds are highly overrated. We need more generalists—people who can provide broad perspectives.” Students are directed to indicate if they agree or disagree and to explain the reasons for their positions. In a similar vein, the second type of writing task (Figure 8) asks students to evaluate an argument (CLA, 2005).

The CLA performance tasks present real-life problems to students such as that for Dyna-Tech and Crime (Figures 9 and 10) by providing an “in-basket” (or nowadays, “computer basket”) of information bearing on the problem (CLA, 2005). Some of the information is relevant, some not; part of the problem is for the students to decide what information to use and what to ignore. Students

A well-respected professional journal with a readership that includes elementary school principals recently published the results of a two-year study on childhood obesity. (Obese individuals are usually considered to be those who are 20 percent above their recommended weight for height and age.) This study sampled 50 schoolchildren, ages 5–11, from Smith Elementary School. A fast food restaurant opened near the school just before the study began. After two years, students who remained in the sample group were more likely to be overweight—relative to the national average. Based on this study, the principal of Jones Elementary School decided to confront her school’s obesity problem by opposing any fast food restaurant openings near her school.

Figure 8. Collegiate Learning Assessment “Evaluate An Argument” Example.
You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech’s sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235. You are provided with the following documentation:

1. Newspaper articles about the accident
2. Federal Accident Report on in-flight breakups in single engine planes
3. Pat’s e-mail to you & Sally’s e-mail to Pat
4. Charts on SwiftAir’s performance characteristics
5. Amateur Pilot article comparing SwiftAir 235 to similar planes
6. Pictures and description of SwiftAir Models 180 and 235

Please prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken into account, and your overall recommendation about whether or not DynaTech should purchase the plane.

integrate these multiple sources of information to arrive at a problem solution, decision, or recommendation. Students respond in a real-life manner by, for example, writing a memorandum to their boss analyzing the pros and cons of alternative solutions and recommending what the company should do. In scoring performance, alternative justifiable solutions to the problem and alternative solution paths are recognized and evaluated.

A closer look at the Crime Performance Assessment provides insight into what a CLA type QR performance assessment might include. Students are posed the problem, provided an in-basket of information, and asked to analyze the information critically and then inform Mayor Stone about their conclusions with evidentiary justification. The in-basket contains the following information:

- Newspaper article about crime in the community
- Research abstracts about drug education program
- Report about success of a drug education program in another community
- Police report (with table of data) about crime and drug use in the community
Pat Stone is running for election as mayor of Jefferson, a city in the state of Columbia. Mayor Stone’s opponent in this contest is Dr. Jamie Eager. Dr. Eager is a member of the Jefferson City Council. You are a consultant to Mayor Stone. Dr. Eager made the following three arguments during a recent TV interview. First, Mayor Stone’s proposal for reducing crime by increasing the number of police officers is a bad idea. “It will only lead to more crime.” Dr. Edgar supported this argument by showing that counties with a large number of policy officers per resident tend to have more crime then those with fewer officers per resident. Second, Dr. Eager said “we should take the money that would have gone to hiring more policy officers and spend it on the XYZ drug treatment program.” He supported this argument by referring to a news release by the Washington Institute for Social Research that describes the effectiveness of the XYZ drug treatment program. Third, Dr. Eager said that because of the strong correlation between drug use and crime in Jefferson, reducing the number of addicts would lower the city’s crime rate. He showed a chart that compared the percentage of drug addicts in a Jefferson zip code area to the number of crimes.

Mayor Stone has asked you to prepare a memo that analyzes the strengths and limitations of each of Dr. Eager’s three main points, including any holes in those arguments. Your memo also should contain points, explain the reasons for your conclusions, and justify those conclusions by referring to the specific documents, data, and statements on which your conclusions are based.

Figure 10. Collegiate Learning Assessment Performance Task (Crime)

- Plots of the relationship between police offers and crime
- Private investigator report about possible connection between opponent and drug education program

In-basket items on a QR assessment might look like those from CLA’s crime task described in Figure 11.

QR and teacher education

In his letter of invitation to this conference, Bernie Madison noted that, “The goal of the conference is to explore educational solutions to the increasing quantitative reasoning demands on US residents, with special focus on the education of teachers.” Apparently one such solution is in hand—add QR to the teacher-preparation and teacher-enhancement agenda. Certainly this can be done. And CLA-type assessments might be used to check to see if teachers have met some expected level of QR—although even if they did, the whole question of pedagogical-content knowledge and classroom practice then needs to be dealt with. That is, can teachers translate their QR into classroom learning
environments where students can acquire the knowledge, skills and abilities that constitute QR?

As may be evident, I am not sanguine about this proposal. The proposal does not get at the heart of the problem—and K–12 teachers are not the cause of the problem. The proposal ignores the current policy and social context of education in the U.S. The policy context is one of high stakes testing. This form of accountability drives what gets taught by teachers in the classrooms. Unless QR becomes a central focus of what is meant by mathematics achievement, and this is very unlikely, it will be put aside even if we accomplished our goals with teachers.

The social context is one of a society that largely does not possess, foster or support QR. That message is broadcast loud and clear, especially to students. The belief about QR goes something like this. QR and mathematics achievement in the U.S. are part of birth—your fixed ability—and some have the right stuff and others do not. (I do not agree with this.) Moreover, the teaching of mathematics K–16—pedagogy, curriculum, context, students—has not met the challenge of creating a quantitatively literate citizenry. Even more strikingly, the existence proof is all around us. In other countries students achieve a much more sophisticated understanding of and ability to do mathematics. International comparisons have made this very clear.
In an article just published in the journal, *Science*, Bloom and Weisberg (2007, p. 996) provide a cogent basis for explaining the U.S. context:

The developmental data suggest that resistance to science [and QR!] will arise in children when scientific claims clash with early emerging, intuitive expectations. This resistance will persist through adulthood if the scientific claims are contested within a society, and it will be especially strong if there is a nonscientific alternative that is rooted in common sense and championed by people who are thought of as reliable and trustworthy.

Since this is a meeting of higher educators, it seems appropriate to lay part of the QR problem at our feet. A study by Liping Ma (1999) highlights the problem. She compared U.S. and Chinese elementary mathematics teachers teaching of mathematics. The U.S. teachers were college graduates; not so the Chinese teachers. Nevertheless, the latter were found to have a strong conceptual grasp of mathematics and emphasized the conceptual in their teaching while the latter had an algorithmic grasp and emphasized algorithmic practice in their teaching. The simple algorithms understood by U.S. teachers were inadequate to the task—e.g., judging when a novel approach taken by a student to a problem was justified. A similar finding, contrasting US and Japanese elementary mathematics teachers, has been reported by Aki Murata (2004). Moreover, in my experience, students who indicate that they will pursue a teaching credential while earning a bachelor’s degree in an academic major may take a somewhat different pattern of courses than “regular” majors. If this pattern is common in mathematics departments, there is then an opportunity-to-learn issue.

If we are not preparing college students adequately in mathematics and quantitative reasoning, perhaps a significant part of the QR problem lies not in teacher preparation but in the preparation of students generally, and teachers in the academic majors. If this chain of reasoning (admittedly not quantitative!) makes any sense, perhaps we should be talking about preparing QR in introductory college mathematics courses for the broad college audience, in general education courses, and in the mathematics major creating a pedagogy that gives the diversity of students access to both QR and the level of mathematics needed to teach in high school.

I understand this is heresy. But perhaps it will stimulate discussion broadly and lead to an analysis of the various contributors to the current state of QR in our country. In the end, we may settle on teacher education. But perhaps if we do, we will have done so with some perspective.
References


