

2011

Communicating Quantitative Literacy: An Examination of Open-Ended Assessment Items in TIMSS, NALS, IALS, and PISA

Karl W. Kosko

University of Michigan - Ann Arbor, kkosko1@kent.edu

Jesse L. M. Wilkins

Virginia Tech, wilkins@vt.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

Kosko, Karl W. and Wilkins, Jesse L. M. (2011) "Communicating Quantitative Literacy: An Examination of Open-Ended Assessment Items in TIMSS, NALS, IALS, and PISA," *Numeracy*: Vol. 4 : Iss. 2 , Article 3.

DOI: <http://dx.doi.org/10.5038/1936-4660.4.2.3>

Available at: <http://scholarcommons.usf.edu/numeracy/vol4/iss2/art3>

Communicating Quantitative Literacy: An Examination of Open-Ended Assessment Items in TIMSS, NALS, IALS, and PISA

Abstract

Quantitative Literacy (QL) has been described as the skill set an individual uses when interacting with the world in a quantitative manner. A necessary component of this interaction is communication. To this end, assessments of QL have included open-ended items as a means of including communicative aspects of QL. The present study sought to examine whether such open-ended items typically measured aspects of quantitative communication, as compared to mathematical communication, or mathematical skills. We focused on public-released items and rubrics from four of the most widely referenced assessments: the Third International Mathematics and Science Study (TIMSS-95); the National Adult Literacy Survey (NALS; now the National Assessment of Adult Literacy, NAAL) in 1985 and 1992, the International Adult Literacy Skills (IALS) beginning in 1994; and the Program for International Student Assessment (PISA) beginning in 2000. We found that open-ended item rubrics in these QL assessments showed a strong tendency to assess answer-only responses. Therefore, while some open-ended items may have required certain levels of quantitative reasoning to find a solution, it is the solution rather than the reasoning that was often assessed.

Keywords

Quantitative Literacy, Assessment, Communication

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

Karl Kosko is currently a postdoctoral fellow for GRIP (Geometry, Reasoning, and Instructional Practices) at the University of Michigan. His primary research interest is mathematical communication with specific focus on the self-regulation of mathematics through mathematical discussion and writing.

Jay Wilkins is Professor of Mathematics Education at Virginia Tech in Blacksburg, Virginia. His major research interests include quantitative literacy, children's development of fractions knowledge, and the teaching and learning of probability and statistics.

Introduction

Over the past thirty years, quantitative literacy (QL) has become a topic of increased focus at both the national and international level (e.g., Cockroft 1982; NCTM 1989; Steen 1999, 2001; Madison and Steen 2008). Sometimes referred to as quantitative reasoning, mathematical literacy, or numeracy, QL focuses not on one's mathematical skills but on the ability to interact with one's world in a quantitative manner (e.g., Steen 1999). An often mentioned, crucial element of such characterizations is communication itself, but there is little to no supporting research that examines the communicative aspect of quantitative literacy such as there is for communication of mathematics (e.g., Lee 2006; Danesi 2007). On the other hand, several well-known large-scale assessments have evaluated QL using closed and open-ended items. These open-ended items can be viewed as opportunities to assess communication in QL. In an effort to examine the state of communication in QL, the current study asks how these large-scale QL assessments evaluate responses to open-ended items for communication.

Background

Communication as a Part of QL

The importance of communication as a part of QL has been described as early as the Cockroft Report (1982), otherwise known as *Mathematics Counts*. Within the pages of this influential report are several references to the importance of communication as part of QL. The report describes that “a numerate person should be expected to be able to appreciate and understand some of the ways in which mathematics can be used as a means of communication...” (p. 11). Although communication was not identified as being the most important component of QL, it was identified as an essential aspect and one that should not be abandoned in pursuit of enhancing numeracy through mathematical skill alone.

The Cockroft Report was followed by other documents in the United States. One of the most influential for mathematics curriculum in the U.S. was the *Curriculum and Evaluation Standards for School Mathematics* published by the National Council of Teachers of Mathematics (NCTM 1989). The 1989 standards document purposefully discussed the concept of QL and identified five goals or requirements for students to obtain a degree of mathematical power in terms of being numerate. Among these was learning to communicate mathematically. Within the same year that the 1989 NCTM standards were released, the National Research Council (NRC 1989) published the report titled *Everybody Counts*. This influential report described QL as a form of number sense evolving from “concrete experience and [taking] shape in oral, written, and symbolic

expression” (p. 47). The report further emphasized the need for problem-solving approaches to be described orally and in writing. Similar to *Mathematics Counts, Everybody Counts* identified that “developing number sense will move children beyond [the] narrow concern for school-certified algorithms for arithmetic” (p. 47), indicating QL should not be concerned solely with arithmetic or calculations.

As the concept of QL began to be popularized, additional literature continued to emphasize the importance of communication in QL. Cobb (1997) characterized QL as requiring “a difficult integration of four very different kinds of thinking” (p. 76), of which communication was one. Contributing to the description by Cobb, Lynn Arthur Steen wrote that teachers “must encourage students to see and use mathematics in everything they do: measurement in science, logic and reasoning in language and communication, ratios and rhythms in music, geometry in art, scoring and ranking in athletics” (Steen 1999, p. 12). Additional literature discussing QL also describes the importance of communication (e.g., Dartmouth College 2009; De Lange 2003; Dingwall 2000; Grawe and Rutz 2009).

It should be noted that the literature thus described emphasizes communication as an essential component of QL and does not seek to claim it as a determining factor. In this line, we do not seek to make any such claims in the current study. Any language that might suggest this should be viewed merely as our attempt to draw focus on an issue that has seen relatively little explicit focus in the literature.

Testing Communication in QL

“A numerate person should be expected to be able to appreciate and understand some of the ways in which mathematics can be used as a means of communication” (Cockroft 1982, p. 11). Yet one of the main ways that QL appears to be measured is by tests of mathematical skill (Steen 2000). Over the past 30 years, tests have been designed to assess aspects of QL. These tests do identify themselves as examining aspects of QL, but do not limit their assessments to this construct. The present study focuses on four of the most widely referenced assessments. These studies included the Third International Mathematics and Science Study (TIMSS-95), the National Adult Literacy Survey (NALS; now the National Assessment of Adult Literacy, NAAL) in 1985 and 1992, the International Adult Literacy Skills (IALS) beginning in 1994, and the Program for International Student Assessment (PISA) beginning in 2000. Below, we include brief descriptions of each of these assessments in terms of their focus on QL. These descriptions are not meant to be comprehensive, but merely to familiarize the reader with these assessments and how they viewed QL.

In writing the description of the TIMSS 1995 study, Martin (1996) stated that the project “plans to report measures of mathematics and science literacy” (p. 1-11). While TIMSS was primarily concerned with comparing achievement scores

across countries as a way of examining curricular and policy differences, there was considerable time and effort devoted to the concepts of mathematics and science literacy (Orpwood and Garden 1998). TIMSS 1995 used the term mathematics literacy rather than QL, but it should be remembered that the terms have often been used synonymously. Additionally, TIMSS 1995 incorporated aspects of “reasoning and social utility” into their conception of mathematics literacy (Garden and Orpwood 1996), indicating a social component to their definition. Orpwood and Garden (1998) clarify this definition by stating that “translation between graphical/quantitative information including statistical to-and-from natural language statements” (p. 30) was part of the reasoning component of mathematics literacy, of which the reasoning and social dimensions were later combined. The social dimension emphasized, “a criterion in the selection of items was that they involve the sort of mathematics question that could arise in real-life situations and that they be contextualized accordingly” (p. 38). Orpwood and Garden (1998) are clear in their specification of mathematical literacy as being distinct from pure mathematical skill. Integration of contextual factors and consideration of communicative aspects were taken into consideration when formulating TIMSS 1995 conception of mathematical literacy and in the construction of items assessing such literacy.

The National Adult Literacy Survey assessed three basic forms of literacy: prose literacy, document literacy, and quantitative literacy (Kirsch et al. 2001). QL is defined as involving “the knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed materials; for example, balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest from a loan advertisement” (pp. 5–6). However, elements from the prose literacy section could also be seen as compliant with definitions of QL by Steen (1999, 2000). Open-ended tasks were included in NALS to “simulate the kinds of activities that people engage in when they use printed materials” (Kirsch et al. 2001, p. 77). The International Adult Literacy Survey was conducted in 1994 as an international version of NALS (see NCES 2010a). As such, its definition of QL is identical while its items used to assess QL appear to be different. De Lange (2003) cited the International Life Skills Study (2000), a later version of IALS, as characterizing QL as partly being communication, therefore extending the NALS conceptions of using open-ended items.

The purpose of PISA might be summed up in the introductory questions of the 2002 technical report. “Are students well prepared to meet the challenges of the future? Are they able to analyse, reason, and communicate their ideas effectively? Do they have the capacity to continue learning throughout life?” (p. 3). Using the term “mathematical literacy,” Adams and Wu (2002) describe PISA’s view of QL as involving the interrelationship of mathematics embedded in

the real world and the students' ability to interpret and use that mathematics usefully. In order to assess the level of mathematical literacy, PISA examined two elements in item difficulty. One is the kind and level of mathematical skill required while the other is the kind and degree of interpretation and reflection required. In examining the role of communication in this form of literacy, PISA described that, at the highest level of proficiency, students are expected to be able to explain or communicate results through use of argumentation (Adams and Wu 2002). These facets of how PISA examined and viewed QL suggest that communication was considered both an aspect of QL as well as something the assessment sought to detect. Further extending this linkage, PISA characterized communication as “expressing oneself, in a variety of ways, on matters with a mathematical content, in oral as well as in written form; and understanding others' written or oral statements about such matters” (OECD 2002, p. 83). In addition to this definition, De Lange (2006) referenced the PISA study as being “concerned with the capacities of students to analyse, reason, and communicate ideas effectively as they pose, formulate, solve and interpret mathematics in a variety of situations” (p. 15). Taken together, it appears PISA defined QL with communication as part of that definition.

The four assessments described here clearly identified themselves as measuring elements of QL. Some even explicitly identified aspects of communication as part of their definition of QL, while others limit the role of communication to the incorporation of open-ended items. However, by including open-ended items, these assessments demonstrated their consideration of communication in QL, even if it was not explicitly stated.

From Mathematical to Quantitative Communication

While communication is described in much of the QL literature, it is our opinion that a sufficient and specific definition for communication in QL is either not present or not widely used. Therefore, it is logical to examine aspects of *mathematical communication*, in which there is a sufficient body of literature, and see how such aspects can be applied to examine communicating quantitatively.

In mathematical communication, the purpose is to communicate about mathematics itself (Lee 2006). In other words, when we engage in the act of mathematical communication, we do so specifically for the purpose of communicating aspects of mathematics in a manner that is related to that purpose. For QL, the purpose of communicating is inherently tied to its context and is not done specifically for or about mathematics (Steen 2001). A newspaper article may use statistical polling data to explain why one candidate is winning an election, but the article is written to communicate about the election, and not the mathematics itself. The margin of error may be discussed to identify the race as a statistical “dead heat” and that there is no clear leader. Yet, while the discussion

may potentially be laden with many quantitative terms and mathematical statements, the primary purpose behind such communication is to describe the nature of the political race.

Contrary to the foregoing example with the newspaper article using the political poll, two statisticians may engage in a discussion about the poll for very different reasons. One might conceive of these two individuals discussing sample size and the specific means used for calculating the confidence intervals and margin of error. While the two statisticians may use the same poll as a basis for argument, in this context their purpose for communicating is inherently different in that it has a distinctive mathematical component. The discussion is wholly mathematical and lacks any substantive elements of other contexts. The scenario with the newspaper and the scenario with the statisticians both use mathematics as the means of communication. However, what might be termed *quantitative communication* involves mathematical means but varying, context-based purposes for communicating about mathematical or quantitative information. Therefore, the difference between mathematical and quantitative communication is the purpose for the communication. Since there is currently no literature describing quantitative communication, it is necessary to identify aspects of mathematical communication that may be present in quantitative communication.

Mathematical communication has been described as being written, spoken, or visually represented (Danesi 2007). Indeed, these different forms of mathematical communication are sometimes used interchangeably or in place of one another (Steele and Johanning 2004). As these are mathematical means of communicating, it is logical that they are aspects of quantitative communication. Therefore, our investigation will examine how various assessments of QL have evaluated written, spoken, and visually represented communication. By *spoken and written communication*, we mean the use of the acts of writing and speech to communicate mathematically either in narrative English, symbolic expression or a combination of the two. By *represented communication*, or *representation*, we mean visual representation such as graphs, charts, diagrams, or other similar means of representing mathematics.

Research conducted on mathematical writing has identified simplistic, procedural, and conceptual types of writing as areas of focus for examination (Kosko, Wilkins, and Pitts Bannister 2009; Shield and Galbraith 1998). While Kosko et al. and Shield and Galbraith used the term descriptive rather than conceptual, the two terms are arguably related. *Simplistic communication* consists of memorized statements or the answer to a problem/task (Kosko et al. 2009). *Procedural communication* entails descriptions of procedures or strategies while *conceptual communication* consists of explanations, justifications, or conjectures (Schleppegrell 2007). Visual representations have not been identified as being procedural or conceptual, but it is logical to conclude that they could be used in

both procedural and conceptual ways, but may only be determined as such through use of either written or spoken communication. Dossey's (1997a) description of measuring QL also uses the terms procedural and conceptual in referring to the different levels an individual engages quantitatively with the world.

Given this background in mathematical communication, it is prudent to investigate how the different large-scale assessments measuring QL (i.e. NALS, IALS, TIMSS-95, PISA-2006) have examined the different forms of communication. Therefore, the current study seeks to answer the following research questions for quantitative communication:

- To what degree do assessment rubrics in large-scale examinations of QL examine responses that are simplistic, procedural, conceptual, and/or include a representation?
- Do item-rubrics assessing mathematical communication (mathematical purpose) differ from those assessing quantitative communication (context-based or quantitative purpose) in the degree to which the rubrics of such items assess simplistic, procedural, conceptual, and/or representation responses?

Methods

Sample

Public-released items and rubrics from the IALS, 2006 PISA study, 1995 TIMSS study, and 1985/1992 NALS (IEA 1995; NCES 2010b, c; OECD 2010) were used as the sample for the current study. These four studies claimed to assess QL / numeracy and each included open-ended items. Therefore, we examined the public-released items for each assessment, along with each item's rubric. Counts of these items are listed in Table 1. We included document and prose literacy items within the NALS 1985/1992 counts based on Dossey's (1997b) descriptions of those items as assessing forms of QL.

Table 1

Numbers of Public-Released Items for the Large-Scale Assessments of Quantitative Literacy.

QL Assessment	Total Items	Open-Ended Items
IALS	15	15
2006 PISA	89	64
1995 TIMSS	23	12
1985/1992 NALS	110	36
Total Items Examined	237	127

Measures

Form of Communication. We examined whether the communication assessed by the large-scale QL tests involved simplistic, procedural, or conceptual communication, or asked for representations (*simplistic* = 1; *procedural* = 2; *conceptual* = 3; *representation* = 4). This coding scheme was adapted from previous studies on mathematical writing (Kosko et al. 2009; Shield and Galbraith 1998). The codes were assigned so that every item that contained any of these codes was assigned each. Therefore, an item rubric may have been coded 1,2,4 to denote that it looked for simplistic and procedural communication and looked for the creation or extension of a representation as well. This was done to examine the combination of mathematical communications assessed instead of simply the most sophisticated or a particular type.

Simplistic communication was deemed to be a focus of an assessment when an answer was sought by the scorer. This included filling in a blank as well as open spaces for writing or coded oral responses. Procedural communication was coded when the rubric gave credit for a description of procedures or strategies. This did not include showing one's work, but needed to be a written account of the procedures. Conceptual communication was coded when the rubric gave credit for descriptions, explanations, justifications, or conjectures. Representation was assigned when the rubric gave credit for extension or creation of mathematical representations included in the answer, such as constructing or labeling a graph.

Figure 1 illustrates an example item and its rubric. While the item asks test-takers to explain how they would solve the problem, the form of communication was assessed from the item rubric, which clearly is looking for a procedural statement since it seeks an explanation of "the basic steps in computing." By looking for an explanation of steps in computation, the rubric seeks a description of the procedure used. All items from all assessments were coded in a similar manner.

Purpose of Communication. A separate code was used to assess the purpose of communications in open-ended questions. A *mathematical purpose* was assigned if the assessment sought out a reply for a specifically mathematical context. A *quantitative purpose* was assigned if the assessment sought out a reply for a context other than, but possibly related to, mathematics. Additionally, a *loose quantitative purpose* was assigned if the assessment's context appeared contrived. Brief examples of items with each form of communicative purpose are shown in Figure 2. One notable feature of items such as the one presented in Figure 2 is that the purpose of the item (e.g., mathematical vs. quantitative) does not necessarily determine the response type (e.g., simplistic, procedural, conceptual) and vice versa. These examples are provided here for descriptive purposes alone, and such

conjectures about the relationship are reserved for analysis of the data and interpretations of such results.

Item from NALS 1992:

You need to borrow \$10,000. Find the ad for Home Equity Loans in the newspaper provided. Explain to the interviewer how you would compute the total amount of interest charges you would pay under this loan plan. Please tell the interviewer when you are ready to begin.

FIXED RATE • FIXED TERM

HOME EQUITY LOANS 14.25%

Annual Percentage Rate
Ten Year Term

SAMPLE MONTHLY REPAYMENT SCHEDULE

Amount Financed	Monthly Payment
\$10,000	\$156.77
\$25,000	\$391.93
\$40,000	\$627.09

120 Months 14.25% APR

Rubric for Correct Answer:

The answer is correct if the respondent explains the two basic steps in computing the total interest charges. The two basic steps are:

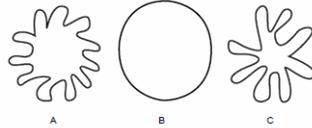
1. The monthly payment (\$156.77) times the number of payments (120) equals the total loan payment.
2. The total loan payment minus the amount of the loan (\$10,000) equals the total interest charges.

Also acceptable is an answer where the respondent explains one but not both steps in computing the total interest charges or is vague about the steps (for example, stating that one needs to know how much one pays over 10 years). Scores distinguished between answers that explained the two basic steps and those that explained only one step or gave less detail.

Figure 1. A 1992 NALS item and rubric that assesses procedural communication. (NCES 2010b).

Mathematical Purpose:

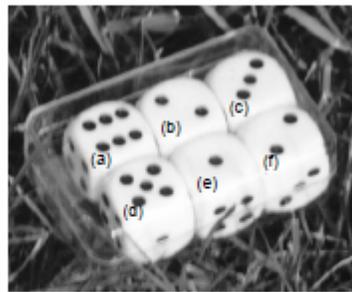
Which of the figures has the largest area? Explain your reasoning.



Loose Quantitative Purpose:

In this photograph you see six dice, labelled (a) to (f). For all dice there is a rule:

The total number of dots on two opposite faces of each die is always seven.



Write in each box the number of dots on the **bottom** face of the dice corresponding to the photograph.

(a)	(b)	(c)
(d)	(e)	(f)

Quantitative Purpose:



Researchers have found out that an ideal coin system meets the following requirements:

- diameters of coins should not be smaller than 15 mm and not be larger than 45 mm.
- given a coin, the diameter of the next coin must be at least 30% larger.
- the minting machinery can only produce coins with diameters of a whole number of millimetres (e.g. 17 mm is allowed, 17.3 mm is not).

You are asked to design a set of coins that satisfy the above requirements. You should start with a 15 mm coin and your set should contain as many coins as possible. What would be the diameters of the coins in your set?

Figure 2. Items from PISA that exemplify each purpose type (OECD 2010).

Analysis

The first step in our analysis was to examine all public-released items from each assessment and select all open-ended items in the QL assessments (Table 1, column 3). Second, we examined each item's rubric to see what type(s) of response was being sought (*simplistic; procedural; conceptual; representation*). Third, we judged the purpose of the response sought by examining both the item and the item's rubric. For example, rubrics in which the purpose of the test-taker's response was to communicate explicitly about the mathematics involved was coded as having a mathematical purpose; rubrics whose communication purpose was to communicate about the context, based upon the underlying mathematics, was coded as having a quantitative purpose. Finally, after all rubrics were coded, we looked for: (1) relationships between the different response types (*simplistic, procedural, conceptual, representation*); and (2) differences between the QL tests (e.g. IALS, PISA) in terms of response type and purpose.

Results

General

We found that 78.7% of the item rubrics sought a simplistic response; 6.3% sought a procedural response; 15.0% sought a conceptual response; and 9.4% sought a representation.

Only 8.8% of the item rubrics called for multiple response types. The other 91.2% broke down as follows: 70.0% of the item rubrics sought only a simplistic response; 3.9% sought only a procedural response; 10.2% sought only a conceptual response; and 7.1% sought only a representation.

Relationships between Different Response Types

Tables 2 and 3 show the six two-way comparisons between response types across all items. As can be seen in the comparison between simplistic and procedural response types (Table 2), rubrics that sought a simplistic response did not typically seek a procedural one (3%, or 3 out of 100), and vice versa (37.5%, or 3 out of 8). A similar relationship can be observed between simplistic and conceptual response types and simplistic and representation response types (Table 2).

Comparisons between the other pairs of response types yielded similar patterns (see Table 3). For example, no items seeking a conceptual response sought a representation. Also, no item that sought a conceptual response sought a procedural one. Additionally, only one of eight rubrics that sought a procedural response also sought a representation. These patterns observable in Tables 2 and 3

suggest that of the 8.8% of the rubrics that sought a combination of response types, almost every one of these rubrics combined a simplistic response with some other format. In other words, other forms of presenting information outside of simply providing the answer were typically not sought in assessing open-ended items.

Table 2
Comparisons of Response Types: Two-Way Comparisons Including the Simplistic Type

Simplistic vs. Procedural			
	Not Procedural	Procedural	Total
Not Simplistic	22	5	27
Simplistic	97	3	100
Total	119	8	127
Simplistic vs. Conceptual			
	Not Conceptual	Conceptual	Total
Not Simplistic	14	13	27
Simplistic	94	6	100
Total	108	19	127
Simplistic vs. Representation			
	Not Representation	Representation	Total
Not Simplistic	18	9	27
Simplistic	97	3	100
Total	115	12	127

Table 3
Comparisons of Response Types: Two-Way Comparisons Not Including the Simplistic Type

Procedural vs. Conceptual			
	Not Conceptual	Conceptual	Total
Not Procedural	100	19	119
Procedural	8	0	8
Total	108	19	127
Procedural vs. Representation			
	Not Representation	Representation	Total
Not Procedural	108	11	119
Procedural	7	1	8
Total	115	12	127
Conceptual vs. Representation			
	Not Representation	Representation	Total
Not Conceptual	96	12	108
Conceptual	19	0	19
Total	115	12	127

Differences in QL Tests for Response and Purpose Types

Comparison of communication purpose across the four tests is depicted in Table 4. IALS had relatively more rubrics assessing a loose quantitative purpose than other tests (60% compared to 0%, 25% and 18% for NALS, PISA and TIMMS, respectively). NALS contained no rubrics assessing for either a loose quantitative purpose or a purely mathematical purpose. Its entire set of open-ended item rubrics assessed a purely quantitative purpose. PISA was the only test containing items with a mathematical purpose (12.5%). These findings suggest that in terms of the purpose of communication, NALS items were more oriented towards eliciting a response with a quantitative purpose than any other test. Overall, the trends observable in Table 4 indicate that while IALS was more focused on assessing items with a loosely quantitative purpose, the other three QL tests tended to favor items with a solid quantitative purpose.

Table 4
Categorization of Assessments for Quantitative Literacy Open-Ended Items: Purpose.

	Mathematical Purpose	Loose Quantitative Purpose	Quantitative Purpose	Total
IALS	0	9	6	15
NALS	0	0	37	37
PISA	8	16	40	64
TIMSS	0	2	9	11
Total	8	27	92	127

Counts of category of response type sought by the open-ended questions in each test are in Table 5. With the exception of IALS, each test contained items assessing all forms of quantitative communication. However, every test also tended to assess simplistic responses in their open-ended questions more often than any other form of response. While certain tests assessed more of one response type than others, proportionally, the tests were fairly similar in the type of response they sought to assess with their open-ended items. The exception is TIMSS, where less than half of the open-ended item rubrics sought a simplistic response. IALS, NALS, and PISA sought simplistic responses on more than three-quarters of their open-ended items. These results suggest that while TIMSS had relatively fewer rubrics assessing simplistic responses than other tests, for the most part, the different tests were found to be similar in regards to response type.

Table 5
Categorization of Assessments for Quantitative Literacy Open-Ended
Items: Response Type

	Simplistic	Procedural	Conceptual	Representation	Total
IALS	14	0	2	0	15
NALS	30	1	2	5	37
PISA	51	5	11	6	64
TIMSS	5	2	4	1	11
Total	100	8	19	12	127

Discussion

The analysis of open-ended item rubrics for QL assessments illustrated two things. First, with the exception of TIMSS, all tests' open-ended rubrics were more oriented towards assessing simplistic responses than other forms of response. Second, item rubrics assessing simplistic responses seldom looked for procedural or conceptual descriptions, or the use of representation, and, further, the relatively few items with rubrics assessing more than one form of response almost always had simplistic responses as one of those forms. Rephrasing these two points, the majority of open-ended rubrics looked for answers and did not look for representations, procedural descriptions, or conceptual descriptions that might accompany them. While many items on these QL assessments may have required test-takers to think critically and deeply about the mathematics they were using, few items in these tests *assessed* such thinking. Rather, the tests assessed the answer provided, not the reasoning or the ability to communicate such reasoning. In other words, even though we might argue that certain open-ended items require certain reasoning to procure the correct answer, if we do not assess the reasoning but only assess the answer, we have no certain evidence for our claim. With these considerations in mind, the majority of QL assessments appear to assess quantitative communication as the ability to produce a simple answer, which may or may not be associated with showing one's work. Too seldom do these assessments examine whether test-takers can describe what they are doing or why they are doing it.

Wilkins (2010) stated that "although several national and international projects have considered the assessment of quantitative literacy, ultimately the notion of literacy that has been emphasized is one focused solely on mathematical achievement" (p. 286). The results of our examination of rubrics for open-ended items presented here support Wilkins' statement. Although his focus on QL was not on communication, but on the incorporation of beliefs and dispositions, our

study also supports his statement of the underlying issue, that “evaluation of learning is often reduced to the use of measures of achievement alone, which only takes into account one component of the overall quantitative literacy construct” (p. 286).

Recall that early reports focusing on QL, such as the Cockroft Report and *Everybody Counts*, cautioned that examination of QL should go beyond computation or arithmetic. Considering that 70% of open-ended rubrics examined in our study sought only simplistic responses, it appears that many assessments of QL may not be examining QL in its fullest extent. Only 6% of items looked for procedural responses, 15% conceptual, and 9% sought a representation. Recalling that these were open-ended items, and not multiple-choice/response items, these results are initially disturbing. However, one must keep in mind that many open-ended items are necessarily focused on obtaining an answer-only response. This does not mean such items are bad or unnecessary. Nor do we wish the reader to draw such conclusions from this study. Rather, what we wish to draw attention to is the apparent lack of what may be considered sufficient items examining quantitative communication. Across 127 public-released open-ended items from four different QL assessments, only 38 (29.9%) asked for more than just an answer. The vast majority of these came from PISA and TIMSS (Table 5), and 17 of them did not have a wholly quantitative purpose. The question to be asked is whether this is an acceptable percentage of quantitative communication items asking for more than an answer. Further, we do not simply wish to have measures of quantitative communication, but quality measures of quantitative communication.

Defining what it means for test-takers to be assessed as mathematically literate, PISA identified that those at the higher end would, among other abilities, be able to “display other higher-order cognitive processes such as generalisation, reasoning and argumentation to explain and communicate results” (OECD 2002, p. 87). As seen in Table 5, PISA had the highest number of public-released items where the rubrics sought more than an answer-only response. However, it also had the largest number of answer-only open-ended items. Therefore, a question to be asked of PISA, TIMSS and other QL assessments is how many items that seek more than a simple answer are needed to examine test-takers’ quantitative communication; what aspects of quantitative communication should be examined; and in what proportion should these various forms be examined. The results of the present study have demonstrated that, when examining these aspects in terms of response type and purpose type, the proportion of items assessing quantitative communication is reduced dramatically. What happens if we further divide items by how likely they are to effectively assess QL? At such a point, how many of these open-ended items will assess quantitative communication as a reliable indicator of QL?

Conclusion

The implications of our findings are straightforward. If communication is to be regarded as a critical and essential element of what makes an individual quantitatively literate, then quantitative communication must be assessed in QL assessments. To do this, more than a simple answer should be required for a larger number of open-ended items, and an appropriate purpose for such communications should be kept in mind. While many items are, necessarily, answer-only, we believe that far too many require no more than showing one's work and stating the answer. Additionally, a number of items have a questionable commitment to assessing QL since their context exhibits only a loose quantitative connection. Further, the quality of quantitative communication items must be ensured and examined in future research. QL assessments must evaluate whether an individual can reason and argue in a way that demonstrates their QL. If QL assessments do not do this, then they truly are reduced to simply another achievement test that happens to have open-ended items.

Acknowledgment

This study was supported in part by National Science Foundation Grant No. 0737455. Any conclusions stated here are those of the authors and do not necessarily reflect the position of the National Science Foundation.

References

- Adams, R., and M. Wu (eds.). 2002. *PISA 2000 technical report*. Organisation for Economic Co-Operation and Development: Paris, France.
- Cobb, G. W. 1997. Mere literacy is not enough. In *Why numbers count* ed. L. A. Steen, 75–90. New York: College Entrance Examination Board.
- Cockcroft, W. H. 1982. *Mathematics counts*. London: Her Majesty's Stationary Office.
- Danesi, M. 2007. A conceptual metaphor framework for the teaching of mathematics. *Studies in Philosophy and Education*, 26 (3), 225–236. <http://dx.doi.org/10.1007/s11217-007-9035-5>.
- Dartmouth College. 2009. Mathematics across the curriculum: Goals of MATC. Retrieved from <http://www.math.dartmouth.edu/~matc/goals.html>
- De Lange, J. 2003. Mathematics for literacy. In *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges* ed. B. L. Madison and L. A. Steen, 75–89. Princeton, NJ: National Council on Education and the Disciplines. http://www.maa.org/ql/pgs75_89.pdf, accessed September 1, 2009.

- . 2006. Mathematical literacy for living from OECD-PISA perspective. *Tsukuba Journal of Educational Study in Mathematics* 25, 31–35.
- Dingwall, J. 2000. Improving numeracy in Canada. Report prepared for the National Literacy Secretariat. Retrieved September 1, 2009 from <http://www.hrsdc.gc.ca/eng/hip/lld/nls/publications/b/improve-b.pdf>
- Dossey, J. A. 1997a. Defining and measuring quantitative literacy. In *Why numbers count: Quantitative literacy for tomorrow's America* ed. L. A. Steen, 173–186. New York: College Entrance Examination Board.
- . 1997b. National indicators of quantitative literacy. In L. A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* ed. L. A. Steen, 45–59. New York: College Entrance Examination Board.
- Garden, R. A., and G. Orpwood. 1996. Development of the TIMSS Achievement Tests. In *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development* ed. M. O. Martin and D. L. Kelly, 2-1 – 2-20. Chestnut Hill, MA: Boston College.
- Grawe, N. D., and C. A. Rutz. 2009. Integration with writing programs: A strategy for quantitative reasoning program development. *Numeracy* 2 (2), 1–18. <http://dx.doi.org/10.5038/1936-4660.2.2.2>.
- IEA. See International Education Association
- International Education Association. 1995. *Released mathematics and science literacy items population 3*. Retrieved from: <http://timss.bc.edu/timss1995i/TIMSSPDF/CitemMSL.pdf>
- Kirsch, I., K. Yamamoto, N. Norris, D. Rock, A. Jungeblut, and P. O'Reilly. 2001. *Technical report and data file user's manual for the 1992 National Adult Literacy Survey*. Washington D.C.: National Center for Education Statistics: Washington D.C.
- Kosko, K. W., J. L. M. Wilkins, and V. R. Pitts Bannister. 2009. Writing sophistication in students' answers to algebraic questions. *The MathMate* 33 (1), 18–22.
- Lee, C. 2006. *Language for learning mathematics: Assessment for learning in practice*. New York: Open University Press.
- Madison, B. L. and L. A. Steen. 2008. Evolution of numeracy and the National Numeracy Network. *Numeracy*, 1 (2), article 2. <http://dx.doi.org/10.5038/1936-4660.1.1.2>.
- Martin, M.O. 1996. Third International Mathematics and Science Study: An Overview. In *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development* ed. M. O. Martin and D. L. Kelly. Chestnut Hill, MA: Boston College.
- National Center for Education Statistics. 2010a. *Adult literacy and lifeskills survey (ALL)*. Retrieved from <http://nces.ed.gov/surveys/all/>

- . 2010b. *National assessment of adult literacy (NAAL): Sample questions search: 1985, 1992 & 2003*. Retrieved from:
http://nces.ed.gov/naal/sample_items.asp
- . 2010c. *Adult literacy and lifeskills survey: Sample items*. Retrieved from:
<http://nces.ed.gov/surveys/all/items.asp?sub=yes>
- National Council of Teachers of Mathematics. 1989. *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Research Council. 1989. *Everybody counts: A report to the nation on the future of mathematics education*. Washington, D.C.: National Academy Press.
- NCES. See National Center for Education Statistics.
- NCTM. See National Council of Teachers of Mathematics
- NRC. See National Research Council
- OECD. See Organisation for Economic Co-Operation and Development
- Organisation for Economic Co-Operation and Development. 2002. *Programme for international student assessment: Sample tasks from the PISA 2000 assessment of reading, mathematical and scientific literacy*. Retrieved from:
<http://www.oecd.org/dataoecd/44/62/33692744.pdf>
- . 2010. *PISA released items – Mathematics*. Retrieved from:
<http://www.oecd.org/dataoecd/14/10/38709418.pdf>
- Orpwood, G., and R. A. Garden. 1998. *Assessing mathematics and science literacy*. Vancouver, Canada: Pacific Educational Press.
- Schleppegrell, M. J. 2007. The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, 23(2), 139–159.
<http://dx.doi.org/10.1080/10573560601158461>
- Shield, M. and P. Galbraith. 1998. The analysis of student expository writing in mathematics. *Educational Studies in Mathematics*, 36, 29–52.
<http://dx.doi.org/10.1023/A:1003109819256>
- Steele, D. F., and D. I. Johanning. 2004. A schematic-theoretic view of problem solving and development of algebraic thinking. *Educational Studies in Mathematics*, 57(1), 65–90.
<http://dx.doi.org/10.1023/B:EDUC.0000047054.90668.f9>
- Steen, L. A. 1999. Numeracy: The new literacy for a data-drenched society. *Educational Leadership*, October, 3–13.
- . 2000. Reading, writing, and numeracy. *Liberal Education*, 86(2), 26–37.
- . (ed.). 2001. *Mathematics and democracy*. The National Council on Education and the Disciplines. Retrieved from
<http://www.maa.org/QL/MATHANDDEMOCRACY.HTML>

Wilkins, J. L.M. 2010. Modeling quantitative literacy. *Educational and Psychological Measurement*, 70(2), 267–290.
<http://dx.doi.org/10.1177/0013164409344506>