A coding scheme for mathematical knowledge embedded in print media*

Pete Johnson

Eastern Connecticut State University,
Willimantic
johnsonp@easternct.edu

While the last decade has seen a great deal of progress in the area of quantitative literacy, there are still issues with the understanding of this term and lack of agreement on what should be included in quantitative literacy courses. This paper describes the first steps in a study of coding the mathematical knowledge embedded in newspapers, with the ultimate goal of helping to clarify what areas should be highlighted in quantitative literacy courses.

Introduction

Throughout history, there has always been a demand for individuals to be quantitatively literate. However, the field of quantitative literacy (hereafter abbreviated QL) has only taken its present form in the United States very recently. In a recent overview of the history of the QL field, Madison and Steen (2008) assert that interest in QL as a discipline at the collegiate level dates to a report prepared by a QL subcommittee of the Mathematical Association of America (MAA) released in the mid-1990's (Sons, 1996). Since this time, a number of developments have taken place in the QL area. In particular, the formation of the National Numeracy Network (NNN) in 2000 and the Special Interest Group of the MAA on QL (SIGMAA-QL) in 2004 demonstrate professional interest in QL as a discipline.

Despite the interest in QL in the United States, there is still a great deal of disagreement on the parameters of QL. The term "quantitative literacy" has been taken to mean everything from the arithmetical understandings usually obtained in elementary school to a rich, contextual understanding of numerical evidence, utilizing tools from fields such as statistics and pre-calculus. Madison and Steen (2008) note that this disagreement has led to a lack of progress in the area: "Many mathematicians, for example, view numeracy as inherently elementary and thus the responsibility of K-12 schools rather than colleges. Discussions at a 2004 international workshop, *Numeracy and Beyond* … were hampered greatly because of these conflicting views of the nature and complexity of numeracy" (Madison & Steen, 2008, p. 2).

Concomitant with this lack of understanding of the meaning of QL is a lack of agreement on what should be the main foci of the content of QL courses. While there are now a number of published curricula in the QL area (for example, Sevilla &

* This paper was delivered to the15th International Conference of Adults Learning Mathematics-A Research Forum but mistakenly omitted from the conference proceedings.

Somers, 2007), there is no consensus on what content should be included in QL courses or what emphasis should be placed on various topics.

This paper represents my effort to articulate the content on which QL should focus. It is my belief that the focus of QL should be the mathematical content needed to be an informed citizen in our society. This preliminary research study investigates the content of the most widely circulated newspaper in the United States, *USA Today*. In analysing the mathematical content of *USA Today*, I have articulated a number of schemas upon which an informed citizen could draw to understand more fully the news stories that shape our society. It is my hope that future research using this coding scheme could lead to a better consensus of the scope of QL and the content that should be included in future QL curricula.

Methodology

In this study, I am using the newspaper *USA Today* as a starting point. Using this newspaper has several advantages. First, while circulation for newspapers has certainly declined with the availability of television and online news services, it is still true that about 42% of households in the United States receive a daily newspaper (Stepp, 2001). Using a printed newspaper for research provides a permanent record without the worry of electronic links being broken or lost in the future. *USA Today* is the most widely circulated newspaper in the United States, with a daily readership of approximately 3.6 million readers (according to the *USA Today* website). More individuals in the United States thus obtain information on national and international events from this newspaper than from any other single source of print media.

It is important to note that I did not intend for the coding scheme to be merely a count of the numbers or figures used in the newspaper. Instead, I wanted to encode the mathematical knowledge that *could be brought to bear* on a situation described in the newspaper. For example, results from statistical polls have an inherent amount of associated uncertainty. The intent of this coding scheme was to encode this fact, even if a margin of error or other statement of uncertainty was not made explicit in a news article. I also wanted relevant themes to emerge from the data, rather than being imposed at the beginning of the study. I thus utilized the ideas of naturalistic inquiry and emergent design described by Lincoln and Guba (1985).

As analysis proceeded, the data was organized into a number of mathematical and statistical areas that I have called schemas. For purposes of this study, a *schema* is an organizer for some discrete piece of knowledge that could provide a framework for future understanding. A schema includes an identifier and a set of questions that one might ask of some fact or claim made in written text. Examples of these schemas are given in the next section of this paper.

Schemas for Encoding QL Themes

In this section, I have chosen ten different schemas as illustrations. There is no claim that this list is complete or comprehensive; however, based on subjective impressions, these ten schemas do appear frequently in the print media. I have described each schema through a list of questions that an informed citizen should ask of a newspaper article that includes the kind of content described by the schema. Each is also illustrated by a recent newspaper article with an excerpt showing how the schema

applies to the content of the article. The three newspaper articles used as illustrations are as follows:

> "CEO pay climbs despite companies' struggles" (Beck and Fordahl, 2008; hereafter referred to as BF08).
>
> "Childhood cancers more common in NE"(hereafter referred to as CC08).
>
> "Study: Diabetes pills may be enough for many (Marcus, 2008; hereafter referred to as M08).

Statistical schemas

Five of the schemas that emerged from the analysis are statistical in nature. These include ideas involving sampling, variables, experimental design, central tendency, and statistical significance.

Sampling

The vast majority of statistical studies use a sample rather than an entire population. This fact raises the issues of sample selection, sample size, and sample representativeness as it relates to the intended population. An informed citizen using the sampling schema will ask the following questions of a statistical study reported in a newspaper article:

> What is a sample and what is a census? Which is being used in the article?
>
> What is the sample size? Is it sufficient to answer the question(s) being posed?
>
> What is the population to which the study's authors want the results to apply?
>
> Is the sample a reasonable representation of the overall population? Are there any clear ways in which such a sample might be biased in representing the intended population?

The article M08 discussed the need for insulin for patients with diabetes. It was based on a study of 191 veterans, 96 of whom took only oral medication. An informed citizen reading this article and using the sampling schema should be cognizant of the sample size and possibly the representativeness of the sample (in that only veterans were included in the study), and interpret the conclusions of the study in light of these potential limitations.

Variables

Every statistical study includes variables, and these variables need to be operationalized in an appropriate way. An informed citizen using the variables schema will ask the following questions of a statistical study reported in a newspaper article:

- What variables are being considered in the article? What variables are being left out?
- How are the variables measured? Are the variables being measured in reliable and/or valid ways?
- Are there differences in reporting that change the values of the variables?

In the study reported in article M08, the subjects were placed into groups according to whether or not they had switched from oral medication to insulin over a 15-year period. An informed citizen reading this article should ask whether this is the appropriate way to measure the effectiveness of oral medications as compared to insulin. An informed citizen would also ask what variables are left out of the study that would contribute to the necessity of taking insulin.

Experimental Design

The way a statistical study is designed determines the kinds of conclusions one can draw from it. An informed citizen using the experimental design schema will ask the following questions of a statistical study reported in a newspaper article:

- What is the difference between an observational study and an experimental study? When is each appropriate? What kinds of conclusions can be drawn from each?
- What is the difference between cross-sectional and longitudinal research? What kinds of conclusions can be drawn from each?
- What is the role of randomization in statistics? Is it present in the study reported in the article?

The article M08 reports on an observational study, with no randomization of the subjects (as best as can be determined from the article). As such, the informed citizen should realize that it is difficult to make any causal conclusions from it.

Central Tendency

Measures of central tendency are perhaps the most used, and the most important, single measures in statistics. However, it is important for an informed citizen not simply to know how to compute each one, but also to know what their properties are and when one measure is more appropriate than the other. An informed citizen using the central tendency schema will ask the following questions of a statistical study reported in a newspaper article:

- How do we define mean and median?
- What are the possible distortions of a set of data when using the mean? When using the median?
- When is the mean more appropriate to use than the median, and when is the median more appropriate than the mean?
- Are there other measures of central tendency (e.g., a trimmed mean, a mode) that would better represent the data?

The article BF08 reports on the total compensation for CEO's of 410 of the Fortune 500 companies. The article is consistent in using the median rather than the mean as a measure of central tendency. The informed citizen should always ask which measure is appropriate and what, if anything, is being distorted by one rather than the other. Because a small number of CEO's had very large compensation packages, it is clear that the mean compensation would be larger than the median, and so in this case the median is likely to be a better measure of the compensation for a "typical" CEO. On

the other hand, if one wished to know the total compensation for the subjects in this study, this cannot be determined from the median (but could if the mean were given).

Statistical Significance

Because statistical studies virtually always use samples and not entire populations, they are always subject to statistical fluctuation and error. Statistical significance (which here is understood to include margin of error) is a measure of the uncertainty present in data, and is the statistical way to determine whether some result is meaningful or might be simply due to chance. An informed citizen using the statistical significance schema will ask the following questions of a statistical study reported in a newspaper article:

- What does "margin of error" mean? Is this given in the article? How (un)certain are the results in the article?
- Does the article mention statistical significance? If not, do the results reported appear to be statistically significant?
- Does the article mention practical significance? Is there practical significance to the results?

The article CC08 reports the rate of childhood cancers in various parts of United States. While the title of the article reports that "Childhood cancers [are] more common in [the North-East]," it is unlikely, based on the data reported in the article, that the difference is statistically significant; these rates vary by region from 179 to 159 per one million children. Even if the difference in regions reaches a level of statistical significance (this is not reported in this article), it seems unlikely that there is much practical significance to the results.

Working With Numbers

Three of the schemas that emerged from the analysis involve rates, percentages, and large numbers. I have termed these three schemas "working with numbers."

Rates

Rates are a type of number that most adults see in their daily lives. Nevertheless, there are a number of QL issues around the use of rates. An informed citizen using the rates schema will ask the following questions of a newspaper article:

How does one interpret a rate?

Is the basis for the rate reasonable? Is there a more appropriate scale to report the rate?

The article CC08 reports the cancer rate for children. These rates are reported as the number of cancers per one million children. An informed citizen should be able to understand this rate, realize that childhood cancer is quite rare, and ask if the basis for the rate (cancers per one million children) is reasonable, or if another basis might be easier to interpret.

Percentages

Percentages are a special kind of rate that are also encountered by adults on an almost daily basis. An informed citizen using the percentages schema will ask the following questions of a newspaper article:

What is the meaning of a percentage?

Given two numbers, how do we compute the percentage change between them?

How do we interpret a percentage of more than 100%?

How can we compare percentage increases and decreases? (For example, a "20% increase" followed by a "20% decrease" does not return one to the original value.)

The article BF08 includes a number of percentages, most of them percentage changes in executive compensation from one year to another. An informed citizen reading this article should be able to interpret each percentage change, and be able to convert from a percentage change to an actual dollar change in salary and vice versa.

Large Numbers

When issues concerning something as large as the national economy are reported in newspapers, numbers in the billions and trillions are likely to be encountered. Certainly an informed citizen should be able to understand the size of large numbers that (s)he might encounter, such as million, billion, or trillion. Just as important, however, is the ability to put such large numbers into a reasonable context, in order to "see" such large numbers more readily and understand the scope of the issue under discussion. An informed citizen using the large numbers schema will ask the following questions of a newspaper article:

What is the relative size of large numbers such as million, billion, and trillion?

What is an appropriate context for a given large number, to make it easier to interpret? (For example, an expenditure by the government might be interpreted as "dollars per person"; the pay of the CEO of a corporation might be interpreted as "multiple of the typical employee's salary"; etc.)

The article BF08 reported the total compensation package for CEO's of some of the major corporations in the United States. The largest of these numbers is $83.1 million. An informed citizen would be able to place this number in a meaningful context, such as "this number is about 40 times the total amount of money I expect to make in my working lifetime," or "if the median salary at this corporation is $50,000, this executive is making as much money as about 1600 employees at the median salary."

Other Schemas

The final two schemas that emerged from the analysis did not seem to "cluster" with the others (although it would certainly be reasonable to consider "interpretation of graphs" to be a part of the statistical schemas). These two schemas involve interpretation of graphs and scales of growth.

Interpretation of Graphs

A newspaper like *USA Today* uses a fairly large number of graphs. The reader of this newspaper certainly must be able to interpret the information in those graphs to make sense out of many of the articles in the newspaper. An informed citizen using the interpretation of graphs schema will ask the following questions of a newspaper article:

Is the graph chosen in the article appropriate for the data?

Is the graph drawn accurately?

What is left out of the graph?

What is the source of the information shown in the graph?

The article M08 includes a graph, a pie chart showing the proportion of diabetics who take insulin, oral medication, both, or neither. The informed citizen should be able to look at the graph and see how the various regions of the graph compare in size. The informed citizen would also note the source of the data, given at the bottom of the graphic. The figures include all diabetics, not just the ones included in the study reported in the article.

Scales of Growth

Certainly, the growth of some phenomenon is of vital importance in understanding it. An informed citizen using the scales of growth schema will ask the following questions of a newspaper article:

What does it mean for something to grow at a linear rate? A quadratic rate? A cubic rate? An exponential rate? A logarithmic rate? Etc.

What is the growth rate for the phenomena in the article?

How does the growth rate seen in the article compare with other scales of growth?

The article M08 includes the following quote from a researcher reaction to the study being reported in the article: “In the time that the study was done, we have had an almost logarithmic increase in the availability of different agents with different actions to treat diabetes.” What is meant by a “logarithmic increase”? What is an “almost” logarithmic increase? How would logarithmic growth compare to, for example, linear growth or exponential growth? An informed citizen reading this article would ask such questions (and possibly realize that there is essentially no meaning to the phrase “almost logarithmic increase”).

Next Steps

I want to make clear that the research study reported here is very preliminary. The coding scheme reported is only in its developmental stages. I am making no claim that the scheme presented above is complete or that it will not be revised in the future. While I believe the ten examples above capture a fairly wide range of the topics I would wish to include in a QL course, they are not exhaustive, and other schemas will certainly be added to the overall coding scheme. I welcome any and all suggestions on the scheme and how it might be improved.

What this coding scheme does is give us some basis to analyse the kinds of mathematical knowledge that are important in helping our students to become informed citizens in a democratic society. As with any such scheme, one of the first questions asked of it should relate to the reliability and validity of the scheme. Given some set of newspaper articles, one future direction for this research is to enlist other individuals to learn the scheme and code the articles using it, then find measures of inter-rater reliability based on those codes. In order to establish validity, the scheme should be shared with a wider audience in the QL area who could help establish the content and construct validity of this scheme.

Once some measure of reliability and validity is established, I would like to apply the scheme to a week's worth of newspaper articles. This could help to establish some sort of baseline of the QL demands for one to be an informed citizen. In addition, once the scheme is shown to be psychometrically sound, it could be used to help answer research questions that until now have not been studied. For example, how have QL demands changed over a period of time? How do different newspapers, or different media, compare in their QL demands? These are rather fundamental research questions on QL, which heretofore have been difficult to study without an instrument to do so.

References

Beck, R., & Fordahl, M. (2008, June 16). CEO pay climbs despite companies' struggles. *USA Today*, p. 2B.

Childhood cancers more common in NE. (2008, June 3). *USA Today*, p. 6D.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic Inquiry.* Newbury Park, CA: Sage.

Madison, B.L., & Steen, L.A. (2008). Evolution of numeracy and the National Numeracy Network. *Numeracy 1 (1)*, Article 2. Available at: http://services.bepress.com/numeracy/vol1/iss1/art2.

Marcus, M.B. (2008, June 16). Study: Diabetes pills may be enough for many. *USA Today*, p. 4D.

Sevilla, A., & Somers, K. (2007). *Quantitative Reasoning: Tools for Today's Informed Citizen*. Emeryville, CA: Key College Publishing.

Sons, L., Ed. (1996). *Quantitative Reasoning for College Graduates: A Complement to the Standards.* Washington, DC: Mathematical Association of America.

Stepp, C. (2001). Positive reviews. *American Journalism Review, 23*, 58-67.