Statistical Literacy, Reasoning, and Thinking: A Commentary

Robert C. delMas
University of Minnesota

*Journal of Statistics Education* Volume 10, Number 2 (2002)

Copyright © 2002 by Robert C. delMas, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

**Key Words:** Assessment; Cognitive outcomes; Research.

**Abstract**

Similarities and differences in the articles by Rumsey, Garfield and Chance are summarized. An alternative perspective on the distinction between statistical literacy, reasoning, and thinking is presented. Based on this perspective, an example is provided to illustrate how literacy, reasoning and thinking can be promoted within a single topic of instruction. Additional examples of assessment items based on the perspective are offered. The author concludes with implications for statistics education research that stem from the incorporation of recommendations made by Rumsey, Garfield and Chance into classroom practice.

**1. Introduction**

Each of the papers in this issue tries to identify one of three overarching goals of statistics instruction. As put forward by the authors, these goals represent our intention, as instructors, to develop students’ literacy, reasoning, and thinking in the discipline of statistics. After reading all three papers it is evident that while it is possible to distinguish the three goals (and resultant outcomes) at some levels, there is still considerable overlap in the three domains of instruction. One result of this overlap is that the three authors concur on several ideas and themes. In this commentary I will first address some of the common themes, and then attempt to reconcile the apparent overlap in definitions. While the points of view that I present are my own, they are heavily influenced by the assessment book edited by Gal and Garfield (1997), the literature that is thoroughly cited in the articles by Rumsey, Garfield, and Chance presented in this issue, my personal discussions with these three authors, and the students I encounter in the statistics classroom.

**2. Instruction**
Each author calls for the direct instruction of the particular outcome emphasized in her respective article. We are cautioned to not assume that understanding, reasoning, or thinking will simply come in and of itself without making these objectives explicit to the student. Each author also demands that we not only make our objectives clear, but that we follow through on these objectives by planning instruction to develop these outcomes and assessments that require students to demonstrate their understanding, reasoning, and thinking. This suggests that the statistics instructor must coordinate, or perhaps triangulate, course objectives with instruction and assessment so that one aspect of the course feeds into another. When this is accomplished, meaningful feedback is provided to both the student and the instructor.

The three authors have much to offer towards this aim of triangulating objectives, instruction, and assessment. Perhaps the most obvious contribution is to note that all three goals need to be emphasized and developed. Rumsey, Garfield, and Chance challenge us to teach and assess what we claim to be important. Towards this end, each author provides several examples of both instructional approaches and assessment methods along with citations of references and resources that can be used to develop literacy, reasoning and thinking.

Another instructional theme that emerges from the three papers is that interpretation of statistical information is dependent on context. If a procedure is taught, students should also learn the contexts in which it is applicable and those in which it is not. If this is an objective, instructional activities should require students to select appropriate procedures or to identify the conditions that legitimize the use of a procedure. Similarly, a term or definition should not be taught in isolation. If a goal is to develop students’ understanding of the term “mean” within the context of statistics, instructional activities can be designed to help students discover why the mean is a measure of average, contrast the mean with other measures of central tendency, and demonstrate when and where not to use the mean (e.g., a wrong conclusion is drawn when necessary conditions are not met).

3. The Importance of Assessment

I would like to return for a moment to the triangulation of objectives, instruction, and assessment. It seems to me that assessment often does not receive the same attention as instruction, even though it should have the same prominence. I believe we commit an instructional felony when material or an activity is presented that is related to a course objective, yet the resultant learning is not assessed. I have certainly been guilty of this crime. One reason we may not assess a stated objective is that there simply is not enough room in an exam to include everything covered in a course. This is certainly understandable, although assessment does not have to occur only as a function of a formal, written exam. I will have more to say on this later. Another reason for not assessing a stated objective is that it may be difficult to clearly state the type of behavior that demonstrates achievement of the outcome. In either case, it will prove very disappointing to a student when considerable class time is spent on a topic and the student invests considerable time making sure she understands it.

I will argue that an objective that is not assessed really is not an objective of the course. This is similar to Chance’s "number one mantra" that you "assess what you value". It may be the instructor’s objective to present (or cover) the material or to try out some new activity. The claim that this learning is a goal of instruction, however, seems to be a shallow one unless that learning is assessed. If we cannot find room on an exam, then other means of assessment should be explored. Rumsey, Garfield, and Chance provide us with several alternatives to exam-based assessment. I would like to offer another alternative, which is to use instruction as assessment. My preferred method of instruction is through activities, and all of my activities have an assessment/feedback component. Some of the activities provide automatic feedback and the opportunity for self-correction. The feedback often contradicts students’ responses and prompts
them to ask a neighboring student or call on the instructor for a consultation. While this type of assessment does not produce a score that is entered into a student’s record, it does provide "just in time" feedback that can help a student determine whether he has attained an understanding or needs additional help and information.

Even when feedback is built into an activity, some aspects of an activity may require reflection by the instructor outside of class. In my classes, students know that in-class activities collected for assessment receive a grade (in-class work makes up 15% of the final grade) which provides additional motivation for them to engage in the activities. In these cases, I use a simple scale from 0 to 4 to assign a grade to students’ work, write brief comments, and return the feedback by the next class session. While not as immediate as built-in feedback, students still report that the assessment is timely and useful. I have found that scores from in-class activities are predictive of exam performance. In-class grades can account for 10% or more of the variance beyond that which is accounted for by precollege ability indicators such as high school percentile rank and standardized measures of mathematical and verbal ability. This suggests that students can make up for lower levels of academic preparation by engaging in activities that provide corrective assessment.

As mentioned earlier, one of the major difficulties with designing assessments is to know what it looks like to meet an objective. I want to return to the idea that if you can’t describe the student behavior that meets an objective, then it may not represent a true course objective. My argument may be somewhat circular, but that is partly because I believe that effective teaching requires objectives to be connected to instruction, and instruction to assessment. Clear descriptions of student behavior or examples of behavior that demonstrate an objective provide concrete goals for students. Once the student behavior is described, different instructional experiences that might lead to the goal can be imagined. Therefore, defining the student behavior that exemplifies a learning objective provides the impetus for instructional design. If assessments are then derived from the instructional experiences, students can form valid expectations of how their understanding will be assessed. Assessments tied to objectives through instruction should be both meaningful and useful to students.

4. Separating the Three Learning Outcomes

As instructors of statistics, we may sense that there is a true distinction to be made between literacy, reasoning and thinking as cognitive outcomes. However, as pointed out by all three authors, the distinctions are not clear-cut due to considerable overlap in the domains. Each author cited several definitions for their respective outcome of interest. Often, the definition of one area incorporated abilities from one or both of the others. Garfield especially noted many instances where the terms "reasoning" and "thinking" were used interchangeably in the literature. The inherent overlap appears problematic if the goal is to distinguish the three types of cognitive outcome. However, from an instructional perspective, the overlap suggests that a single instructional activity can have the potential to develop more than one of these outcomes.

For example, Rumsey provides useful suggestions for how we can assess students’ data awareness. In her description she suggests that knowing how data is used to make a decision demonstrates a student’s data awareness and, therefore, a level of statistical literacy. Knowing how to use data implies an understanding of the contexts in which different types of data are useful and the types of decisions that are warranted. If this is the case, knowing how data is used seems to fit well with Chance’s definition of statistical thinking, i.e., knowing how to behave like a statistician. It also appears that a student who demonstrates data awareness also demonstrates statistical reasoning because the student is reasoning with statistical ideas and giving meaning to statistical information.
Together, the three authors provide us with at least two different perspectives on how the three outcomes of instruction are related. If we focus on literacy as the development of basic skills and knowledge that is needed to develop statistical reasoning and thinking ("instruction in the basics" as Rumsey puts it), then a Venn diagram such as the one presented in Figure 1 might be appropriate. This point of view holds that each domain has content that is independent of the other two, while there is some overlap. If this perspective is correct, then we can develop some aspects of one domain independently of the others. At the same time, some instructional activities may develop understanding in two or all three domains.

An alternative perspective is represented by Figure 2. This perspective treats statistical literacy as an all-encompassing goal of instruction. Statistical reasoning and thinking no longer have independent content from literacy. They become subgoals within the development of the statistically competent citizen. There is a lot of merit to this point of view, although it may be beyond the capacity of a first course in statistics to accomplish. Training of a full-fledged, statistically competent citizen may require numerous educational experiences both within and beyond the classroom. It may also be the case that the statistical expert is not just an individual who knows how to "think statistically", but is a person who is fully statistically literate as described by Rumsey.
Both perspectives can account for the perceived overlap between the three domains of instruction. It seems, however, that for just about any outcome that can be described in one domain, there is a companion outcome in one or both of the other domains. Earlier I described how the outcome of data awareness could be seen to represent development of statistical literacy, reasoning, and thinking. I believe that this is the case for almost all topics in statistics. If so, then the diagram in Figure 1 is wanting. Figure 2 does a better job of accounting for the larger overlap across the three domains, although it still may overrepresent the separation of literacy from the other two. Another problem with Figure 2 is that alternative diagrams could be rendered where one of the domains represents the objective that subsumes the others. In advanced courses in statistics, it is not difficult to imagine statistical thinking as the overall goal that encompasses and is supported by a foundation in statistical literacy and reasoning.

I will present another example from my personal experience in an attempt to set up an argument for a perspective that I believe accounts for the arguable overlap. When my understanding of confidence intervals was assessed in a graduate level course, emphasis was placed on selecting correct procedures and performing computations correctly. Even when applied to hypothesis testing I was only asked to "accept" or "reject". It seems to me that the instructors were primarily assessing my statistical literacy (at a graduate level), although I’m sure their intention was to affect my reasoning and thinking. As I furthered my understanding of confidence intervals through my own reading, teaching, exploration through simulations, and discussion with colleagues I developed a better appreciation for the link between confidence intervals and sampling distributions. Further exploration of this connection deepened my understanding of how a statement of ninety-five percent certainty is a statement of probability and how a confidence interval represents a set of possible values for the mean of the true population that generated the sample. If a goal of my graduate-level instruction was to foster this type of understanding, I might have encountered assessment items that attempted to assess reasoning about why I can be ninety-five percent certain or why I can draw a reliable conclusion about a population.

I also recall having to memorize assumptions for various statistical tests and procedures and being required to write them down or select them from a set of options. I don’t recall many items where a research situation was described in some detail and I had to identify the appropriate procedure (or procedures) that applied, determine if a specified procedure was appropriate, or state questions needed to
make such determinations. In other words, most test items did not try to determine my level of statistical thinking. Let me assure you that I did encounter these types of questions on the written preliminary examination. I passed that exam even though my graduate courses did not provide many direct instructional experiences. Somehow I had organized the information and experiences encountered in the courses in a way that allowed me to deal adequately with the demands of the examination. I know, however, that I could answer those questions much better now after some fifteen years of experience with statistical application and instruction. I also believe that some of my experience from the past fifteen years could have been represented in my graduate-level coursework, experience that would have provided better preparation for the challenges of the written preliminary examination and my first encounters with data analysis outside of the classroom.

5. An Alternative Perspective

Reflection on these experiences has caused me to consider a different perspective on how we can distinguish the goals of literacy, reasoning and thinking. As I argued earlier, just about any statistical content can be seen to represent literacy, thinking, or reasoning. The content may be neutral in this respect. What moves us from one of the three domains to another is not so much the content, but, rather, what we ask students to do with the content. I propose that we look to the nature of the task to identify whether instruction promotes literacy, reasoning, or thinking. In the same way, it is the nature of the test item that determines which of the three domains is assessed and possibly allows for more than one domain to be assessed by the same item.

Table 1 lists words that I believe provide orientations that require students to demonstrate or develop understanding in one domain more so than in another. If one goal is to develop students’ basic literacy, then instruction can ask students to identify examples or instances of a term or concept, describe graphs, distributions, and relationships, to rephrase or translate statistical findings, or to interpret the results of a statistical procedure. If, instead, we ask students to explain why or how results were produced (e.g., explain the process that produces the sampling distribution of a statistic; explain how the mean acts as a balancing point; explain why the median is resistant to outliers; explain why a random sample tends to produce a representative sample) or why a conclusion is justified, we are asking students to develop their statistical reasoning. Given Chance’s treatment of statistical thinking I believe it is distinguished from the other domains in that it asks students to apply their basic literacy and reasoning in context. As such, statistical thinking is promoted when instruction challenges students to apply their understanding to real world problems, to critique and evaluate the design and conclusions of studies, or to generalize knowledge obtained from classroom examples to new and somewhat novel situations.

Table 1. Tasks That May Distinguish the Three Instructional Domains

| BASIC LITERACY | REASONING | THINKING |
|----------------|-----------|----------|
| IDENTIFY       | WHY?      | APPLY    |
| DESCRIBE       | HOW?      | CRITIQUE |
| REPHRASE       | EXPLAIN   | EVALUATE |
| TRANSLATE      | (THE PROCESS) | GENERALIZE |
| INTERPRET      |           |          |
This conception of the three outcomes also extends to assessment. If I want to assess students reasoning or thinking, but only ask questions that require students to identify, describe, or rephrase, then I have misunderstood the assessment goal. For example, I might ask the following question after students have studied a unit on confidence intervals.

I. A random sample of 30 freshmen was selected at a University to estimate the average high school percentile rank of freshmen. The average for the sample was found to be 81.7 with a sample standard deviation of 11.45.

Construct a 95% confidence interval for the average high school percentile rank of freshmen at this University. Show all your work to receive full credit.

The correct answer is (77.43, 85.98). Students who successfully complete this item may have a deep understanding of confidence intervals, but all I really know about these students is that they have the basic procedural literacy needed to recall a set of steps, plug values into formulas, and carry out computations.

If I want to assess some aspect of their reasoning, I could use the same context and follow the first item with:

II. What does the 95% confidence interval tell you about the average high school percentile rank of freshmen at this university?

From experience, I can anticipate three types of response to this question. The first type demonstrates misunderstandings about confidence intervals and the process that produces confidence intervals. Here are two prototypical examples.

1. Ninety-five percent of the freshmen have a high school percentile rank between 77.43 and 85.98.
2. There is a 95% chance that the true population mean is between 77.43 and 85.98.

A second type of response may illustrate only that the student has memorized a patented response to the question.

3. I am 95% sure that the true population mean is between 77.43 and 85.98.

The third statement may demonstrate that the student understands that this is the correct way to describe a confidence interval and, possibly, that statements 1 and 2 are incorrect. However, statement 3 does not allow us to make any inference about a student’s reasoning. A third category of responses demonstrates that students have some idea of the process that produces confidence intervals.

4. If we were to draw a random sample of size 30 over and over again, 95% of the confidence intervals would capture the true population mean.
5. There is a 95% chance that this is one of the samples where the true population mean is within the confidence interval.

6. Any population with a mean between 77.43 and 85.98 could have produced the sample.

Knowing that these are the typical types of responses that are given by students, I could write a different type of assessment item to prompt students’ statistical reasoning. Here is a new prompt for the item.

III. Researchers drew a random sample of 30 freshmen registered at a University to estimate the average high school percentile rank of incoming freshmen. The 95% confidence interval constructed from the sample goes from 77.43 to 85.98. Listed below are several ways that the researchers could interpret the confidence interval. For each statement, determine whether or not it is a valid interpretation of the confidence interval. For each statement that is not valid, state why it is an incorrect interpretation of the confidence interval.

The prompt could be followed by all or some subset of the prototypical statements with ample room provided for written responses.

So far, the items I have suggested would allow me to assess some aspects of students’ statistical literacy and reasoning regarding confidence intervals. An additional item or two could build on the problem situation presented above to assess students’ statistical thinking, or their ability to apply their understanding. Here is a suggestion:

IV. A psychology professor at a State College has read the results of the University study. The sociology professor would like to know if students at his college are similar to students at the University with respect to their high school percentile ranks. The sociology professor collects information from all 53 freshmen enrolled this semester in a large section (321 students) of "Introduction to Psychology". Based on this sample, the 95% confidence interval for the average high school percentile rank goes from 69.47 to 75.72.

Below are two possible conclusions that the sociology professor might draw. For each conclusion, state whether it is valid or invalid. For each invalid conclusion, state why it is invalid. Note that it is possible that neither conclusion is valid.

a. The average high school percentile rank of freshman at the State College is lower than the average of freshmen at the University.

b. The average high school percentile rank of freshman enrolled in "Introduction to Psychology" at the State College is lower than the average of freshmen at the University.

In writing the above item, I have in mind students enrolled in an introductory statistics course with an objective that students develop an understanding of assumptions behind statistical inference. The item gives the students a context and offers two interpretations that must be evaluated. Students who understand that a representative sample from the population of interest is needed for comparison should judge statement ‘a’ as invalid. Students who understand that a reliable way to obtain a representative sample is with random selection would judge statement ‘b’ as invalid. However, some students may argue that while the sample was not drawn randomly, it still may be representative of freshmen who take the course. The argument might go that since it is a large section of a general education type course, there might be very little change in student
representation from one term to the next. If it were a representative sample, then statement ‘b’ would be valid. For full credit, I would want a student to also state that the psychology professor needs student information from several semesters to support the claim of a representative sample.

6. Putting it All Together

The above analysis makes certain assumptions about students’ instructional experiences. I argued earlier that objectives, instruction, and assessment should all be interconnected, and I am therefore obligated to offer some description of instructional activities that might prepare students for these assessment items. Joan Garfield, Beth Chance, and I designed an activity to help students confront misinterpretations of confidence intervals, such as "95% of the observed values will be within the limits" or "there is a 95% chance that the population mean is in the interval". The activity and software (delMas 2001) that it is based on can be downloaded for free at www.gen.umn.edu/faculty_staff/delmas/stat_tools/index.htm. The activity takes the students through a series of simulations designed to confront misinterpretations and help them construct correct interpretations of confidence intervals. This type of activity should provide students with some of the background needed to answer problem III. In addition, every time I have students construct a confidence interval, it is always with real data, always within a context, and I always ask them for an interpretation. Students can work together and have them critique each other’s interpretations. I also ask students to share their interpretations with the class and ask if adjustments need to be made in their wording. I often prompt students to determine if the stated interpretation is justified by observations made during the confidence interval activity.

Additional activities would be needed to develop the statistical thinking required by problem IV. I should first note that much of the understandings needed to engage in statistical thinking about confidence intervals might already be in place prior to direct instruction on confidence intervals. If a course already uses real data sets as part of the instruction, then students should have a good sense of data. If students have an opportunity to draw or generate random samples throughout the course, then another important piece is provided. If an activity such as "Random Rectangles" (Scheaffer, Gnanadesikan, Watkins, and Witmer 1996) is used during instruction about sampling, then students have an opportunity to understand the role of random samples. All of these approaches support Rumsey’s call for the development of statistical literacy and point out its role in supporting reasoning and thinking.

Instruction would also need to include direct experiences with judging the validity of conclusions drawn from confidence intervals. An activity might present several scenarios and ask students to judge whether or not it is appropriate to construct a confidence interval given the circumstances (i.e., are all assumptions met?). I would ask students to calculate confidence intervals only for situations that meet the conditions or to not calculate confidence intervals at all. As with Rumsey’s suggestion that we motivate the HOW with the WHY and WHAT, mechanically carrying out procedures is de-emphasized in the activity and the emphasis is placed on decision making. This should help introductory students develop the type of mindset Chance sees as necessary for statistical thinking. Additional scenarios that are very similar to the one presented in problem IV should be used in class, perhaps having students work in small groups to critique them followed by an instructor-led debriefing. Whatever is used, the activities should be tied to the intended assessment, and the assessment tied to the intended outcome.

7. Assessing Practice

For the most part, readers may find that most of what is presented in the three articles and in this commentary makes sense. Many of the recommendations for reform in statistics education are based on sound learning theories and educational research. Many of the recommendations come from leaders in statistics education who speak from their years of experience in the statistics classroom. As such, putting
these recommendations for the development of students’ statistical literacy, reasoning, and thinking into practice really should help to improve the final product of our statistics courses. But does it?

While the body of statistics education research continues to grow, there are still many claims that have not yet been tested. The articles presented in this issue make numerous claims about what should be taught, how it should be taught, and how learning should be assessed. Are these recommendations effective (i.e., do they lead to the intended outcomes)? What are the possible variations for implementing these recommendations and are they all equally effective? Are there other instructional approaches that are just as effective? Does the same approach work for all students and, if not, what are the moderating variables?

The triangulation of objectives, instruction, and assessment within a course can support statistics education research within the classroom. An instructional activity that is thoughtfully designed so that objectives, instruction and assessment are interconnected and provide feedback to each other provides an opportunity that is ripe for inquiry. Most of the hard work that goes into an informative educational research study is already done at this point. With a little imagination the instructor might devise a way for half of the students in a course section to receive the instructional activity while the other half receives "direct instruction" in the topic, perhaps through the use of a workbook or online delivery of content. Because the assessment is already devised and linked to both course objectives and instruction, a meaningful measure of understanding is already in place. In many cases, all that is left is to select the appropriate method for comparison and conduct the analysis. I hope that readers will take the suggestions and perspectives presented in this issue to not only change their instructional approaches, but to engage in practical and meaningful research that can further our understanding of how to best educate statistics students.

References

delMas, R. (2001), *Sampling SIM* (version 5) [Online].
(http://www.gen.umn.edu/faculty_staff/delmas/stat_tools/)

Gal, I., and Garfield, J. B. (eds.) (1997), *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press and the International Statistical Institute.

Scheaffer, R. L., Gnanadesikan, M., Watkins, A., and Witmer, J. A. (1996), *Activity-Based Statistics: Student Guide*, New York: Springer.

Robert C. delMas
General College
University of Minnesota
354 Appleby Hall
128 Pleasant Street SE
Minneapolis, MN 55455
USA
delma001@umn.edu
This article is based on comments made by the author as discussant of a symposium presented at the 2000 Annual Meetings of the American Educational Research Association in New Orleans.