LEARNING GOALS: THE PRIMACY OF STATISTICAL KNOWLEDGE

Nick J. Broers
Maastricht University, The Netherlands
nick.broers@stat.unimaas.nl

Amongst researchers of statistics education and statistics educators alike, statistical literacy, statistical reasoning and statistical thinking have gained prominence as important learning goals for the teaching of statistics. Careful examination of the three concepts shows that considerable disagreement on their definition still exists, creating problems in the attempts to develop valid and useful measurement instruments. It is argued that the fuzziness of the three constructs stems from the fact that their conception was not motivated by empirical regularities in need of explanation, but rather by the desire to create new perspectives on the future development of statistics education. The inherent ambiguity of the three concepts makes them unsuitable as learning goals for statistics education. By focussing on different aspects of statistical knowledge, however, the intended differentiation in meaningful learning goals can be met in a less disputable way.

INTRODUCTION

In recent years, the community of statistics educators seems to have agreed upon three major learning goals for students of statistics. These goals are known as statistical literacy (SL), statistical reasoning (SR) and statistical thinking (ST). Whereas in the bad old days students used to be trained to become proficient in a host of computational skills, modern statistics education makes use of real life data sets in which computation is largely left to specialized computer software. The prime objective is to teach the student to look at the data in an informed way, to understand how graphical displays may be used to detect interesting patterns, to be able to meaningfully interpret summary statistics, and to make use of statistical models intelligently in order to derive new insights. The quality of teaching, as well as the quality of teaching technologies aimed to facilitate learning, can in principle be measured by determining the extent to which these objectives have been met. Or, to rephrase this in currently popular terminology, the quality of our efforts to teach our students statistics can be derived from our success in turning them into statistically literate citizens, proficient in reasoning statistically and manifesting statistical thinking.

Since these terms were first introduced, they have gained importance and prominence, as evidenced by a IASE sponsored website devoted to the development of statistical literacy, recurrent conferences focusing on research related to statistical literacy, and a host of journal articles and books covering research on SL, SR and ST. Notwithstanding its growing popularity, the use of these terms is not without problems.

STATISTICAL LITERACY

Usually, discussions of what it means to be statistically literate stress something like “being able to function as an educated member of society in this age of information” (Rumsey, 2002). Not surprisingly, definitions of SL show a strong focus on the presence of a statistical knowledge (SK) base. Sometimes it seems that in order to be able to distinguish SL from other competencies like SR or ST, definitions stress the existence of primarily elementary knowledge. This is manifest for example, where Garfield (1999) defines SL as “the understanding of statistical language: words, symbols and terms. Being able to interpret graphs and tables,” where Snell (1999) states “the ability to understand statistical concepts and reason at the most basic level.”

As Rumsey (2002) concedes, many authors tend to use the term SL in quite different ways. What is notable on the diverse attempts at defining SL is that such discussions usually take two different turns. On the one hand, an attempt is made to stipulate what it is exactly, that a statistically literate citizen should know of statistics, i.e., what type of SK is required to be able to be called statistically literate. These discussions tend to be clear, although not very consistent. For example, Gal (2000) lists a whole scale of topics that seem to correspond with what we usually consider to be learning goals for most students of tertiary education, ranging from understanding which type of study was used and what sample was drawn to knowledge of graphs and probability statements. Others, like Watson (1997), connect the development of SL to basic understanding of statistical terminology. The other turn that attempts at defining SL take focus on additional requirements for becoming statistically literate; requirements other than SK elements. Here, the discussion of what it means to be statistically literate becomes far less clear. For example, we find reference to the importance of data awareness (Rumsey, 2002), or to "the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private" (Wallman, 1993, p. 1), and even to the assertion that SL "…suggests a broad cluster (….) of desired beliefs, habits of mind, or attitudes, as well as general awareness and a critical perspective" (Gal, 2004, p. 48). These latter definitions stem from the desire to distinguish SL from mere SK.

Gal (2004) in particular, has done much work in trying to give a clear view of the relationship between SL and SK. He has developed a model of SL that shows SK to be only one of several knowledge elements required for becoming statistically literate. The other elements being literacy skills, mathematical knowledge and context knowledge. Apart from these diverse knowledge bases, SL also requires certain beliefs and attitudes, as well as a critical stance. It would therefore seem that SL has only marginally to do with statistics per se.

Taking a critical view of the model proposed by Gal (2004) leads to the obvious question of what you end up with if you subtract the SK from the statistically literate citizen. This residue is in fact a well informed, intelligent and skeptical citizen. Provide such a person with knowledge of statistics and he will deploy this knowledge in an intelligent way and thus demonstrate statistical literacy. However, teach statistics to someone who is not by nature intelligent, skeptical or well informed, and it will be difficult to educate such a person into becoming statistically literate.

STATISTICAL REASONING AND STATISTICAL THINKING

After the elaborate if somewhat vague description of SL, it seems difficult to identify other modes of statistical maturity that do not amount to the same. In fact, as Chance (2002) observes "many appear to use 'thinking,' 'reasoning,' and 'literacy' interchangeably in an effort to distinguish the understanding of statistical concepts from the numerical manipulation that too often has characterized statistical use and instruction." Nonetheless, SR is considered as a learning goal in its own right, to be distinguished from SL. SR is defined as "the way people reason with statistical ideas and make sense of statistical information" (Garfield, 2002). As Garfield adds "this involves making interpretations based on sets of data, graphical representations and statistical summaries." Like SL, SR rests on the availability of a suitable SK base: "underlying this reasoning is a conceptual understanding of important ideas, such as distribution, center, spread, association, uncertainty, randomness and sampling" (Garfield, 2002). In comparison to SL, SR seems much more limited in scope. Basically it involves the way people derive conclusions on the basis of SK. Whereas SL requires the subject to be able to read, to understand numbers, to be intelligent, to be skeptical etc., it would seem that SR only requires logical ability, apart from SK. However, delMas (2002) takes a different view on the distinction between SL and SR. He sees SL and SR hierarchically related, with SL being restricted to basic operations like being able to identify examples of a term or concept, being able to describe graphs, distributions and relationships and being able to interpret the results of a statistical procedure. SR, according to delMas, is more concerned with understanding why or how certain results were produced, like understanding the role of the sampling distribution in the calculation of a confidence interval. So according to delMas, SL is more or less prerequisite to SR, but not the other way around.

Like SL, ST is another term that sounds familiar but is difficult to demarcate in a useful way. Chance (2002) presents a number of attempts to define ST, such as that offered by Snee (1990, p. 116): "thought processes, which recognize that variation is all around us and present in everything we do, all work is a series of interconnected processes, and identifying, characterizing, quantifying, controlling and reducing variation provide opportunities for improvement." Other definitions discussed by Chance (2002) also stress the awareness of the key role of variation in processes.

However, being aware of the omnipresence of variation and the need for modelling uncertainty does not make clear how to operationalize ST as a learning goal. More clarity is provided by the research of Wild and Pfannkuch (1999), who developed an elaborate four dimensional model of ST based on interviews with practising statisticians and students of statistics. ST has been considered by Ullman (1995) as a fundamental intelligence. Studying the model of Wild and Pfannkuch one becomes aware that ST is indeed a fundamental method of enquiry that cannot simply be learned but that has to mature in talented individuals who by nature possess this type of intelligence. Whereas our students are taught how to summarize data graphically and numerically, and how various models may be used to infer probabilistic statements on target populations, the propensity to think statistically cannot easily be transferred in a classroom situation.

One of the dimensions of the ST model clearly contrasts the statistical thinker with the user of statistics who does not display ST. The interrogative cycle, as Wild and Pfannkuch call this dimension, opens with the generation of possible causes, explanations and mechanisms, of possible ways of extracting information from the real system, of formulating plans of approach to the problem. This generation of possibilities will not only come from SK but also from the context. It is the interplay of SK with contextual information that turns the use of statistics into an imaginative enterprise. The phase of generation is followed by the seeking of more information by reading relevant literature, consulting with colleagues and context-matter experts and by collection and querying of data. This is followed by a critical interpretation of the data. Whereas many students tend to jump from data analysis to judgment, the statistical thinker examines whether the suggested interpretation makes sense: "…(we will be)….arguing with ourselves, weighing up against our context knowledge, against our SK, against the constraints we are working under, and we anticipate problems that are consequences of particular choices" (Wild and Pfannkunch, 1999, p. 232). The critical interpretation results in a judgment in which a number of preconceptions will be discarded and we come to a formulation of our present understanding. Wild and Pfannkuch underline that the judgment is not a simple answer to a research question, but involves an evaluation of the reliability of the information, a consideration of rival explanations, a judgment of the extent to which context-matter and statistical understanding agree, etcetera. In line with this model, Chance states that ST implies mental habits like "consideration of how to best obtain meaningful and relevant data to answer the question at hand…(…)…seeing the complete process with constant revision of each component…(…)…omnipresent skepticism about the data obtained…(and above all)…thinking beyond the textbook" (Chance, 2002). ST is the very antithesis of the dreaded cookbook application of statistics. To stimulate the development of ST through statistics instruction, both Chance (2002) and Wild and Pfannkuch (1999) lay stress on the provision of research projects for students, in which the use of statistics is embedded in a research question pertaining to a particular context, and every step in the project demands the interplay between statistics and the problem-context.

MEASURING STATISTICAL LITERACY, REASONING AND THINKING

To be meaningful, it must be possible to assess the extent to which a learning goal has been met. In this case, it should be possible to judge whether a person's degree of SL has improved, whether he is apt at SR or whether he shows an adequate propensity to think statistically. However, as soon as we attempt to measure SL, SR and ST it becomes obvious that the respective definitions are not really that clear cut and that the three domains tend to show considerable overlap (see delMas, 2002). For example, the assessment tasks that Chance (2002) presents for measuring ST seem designed to induce the student to search for context knowledge and to pose critical questions, elements that figure prominently in Gals model of SL (Gal, 2004). Likewise, the 10 "worry questions" that Gal (2004) presents as questions the statistically literate person is supposed to have in mind, all seem to reflect a mix of ST and SR. The confusion arises because SL, SR and ST are postulated and treated as theoretical constructs, but they are not, as least not in the traditional way.

Traditionally, theories start off in response to empirical regularities and our desire to seek exploratory mechanisms that will account for these regularities. Such explanatory mechanisms are often postulated in the form of theoretical constructs, explanatory concepts that are rooted in a theory. The postulation of such constructs is usually followed by careful conceptualization and eventual operationalization in the form of a measurement instrument. When shown to be both reliable and valid, the measurement instrument will then be used for empirical verification of the supposed explanatory role of the construct. Thus, the fact that some people are more apt in understanding and solving a variety of problems than others has in the past lead us to postulate the existence of the theoretical concept of intelligence. Careful conceptualization has since resulted in the operationalization of this construct into a variety of IQ-scales, that can be used to predict the diversity in performance on problem solving tasks. SL, SR and ST were not postulated in response to empirical regularities that needed theoretical explanation. Instead, the three constructs were created in response to the desire to formulate more modern and meaningful learning goals for statistics education. It had long been recognized that the traditional way of teaching statistics, with its heavy emphasis on formal probability and computational skill, did not succeed in realizing the full potential of statistics as a tool for extracting information from a world dominated by uncertainty. A shift in emphasis was deemed necessary (e.g., see Cobb, 1992; Moore, 1992) and in the wake of the search for new and more meaningful learning goals SL, SR and ST were postulated as constructs that could help us in the search for new directions in statistics education; constructs that could help us get a better view of what we should be teaching and how best to teach it. Their existence is not dictated by empirical observations, but rather empirical observations are sought in order to justify their creation. It is the individual researcher who decides what will be included in a definition and what will be left out. Contrary to present efforts, we need not question in what way SL, SR and ST are distinct; rather, we should determine if and to what extent we feel the three terms should be distinguished from each other.

Not until we have firmly made up our minds over what we wish SL, SR and ST to be, can we start to think of suitable assessment procedures or measurement instruments. Of the three constructs, to date only research on SR has resulted in a reliable measurement instrument. The SRA scale, developed by Garfield (2003) contains eight different scales for assessing statistical reasoning, each pertaining to a different subdomain of statistical knowledge (SK). Different scales pertain to the interpretation of probabilities, to the selection of averages, to the interpretation of correlations, etc. Another eight scales measure commonly held misconceptions with regard to these same subdomains. The sixteen items all directly pertain to SK, and probe the proficiency of the student to actively use this knowledge. This makes an assessment of SR less disputable than potential assessments of SL or ST, where many more non-statistical elements like critical stance, awareness of the problem-context etc., presumably need to be captured. The SRA can be used as an instrument to provide an overall score of a students ability to reason statistically, but as Garfield (2003) admits, the validity of this operation is questionable, given that the eight scales show low intercorrelations. Apparently, students who reason correctly on probabilities need not do so with regard to correlations. It seems more reasonable to compute a reasoning score for each subdomain separately, as Garfield has done in a cross cultural study (Garfield, 2003). However, what this shows is that in using this instrument we are not interested in assessing a students overall ability in reasoning with statistics, but we want to know to what end the student is able to make active use of particular SK in a meaningful way. In this respect, it is not clear why the assessment should be restricted to these 8 particular subdomains.

THE PRIMACY OF STATISTICAL KNOWLEDGE

Notwithstanding the advent of popularity of SL, SR and ST as concepts that structure our thinking on what to teach and how to assess its results, traditional assessment of statistics education is still focused on the assessment of surface knowledge. Obviously, more meaningful learning goals are desirable. As has been argued, the concepts of SL, SR and ST are inherently fuzzy and any attempt at unambiguous definition will inevitably contain arbitrary choices that need not meet the approval of fellow researchers. However, all definitions that have been suggested thus far are organized around a core of SK. The three concepts do seem to have a somewhat different bearing on this knowledge base. To show how, it is instructive to make a distinction in types of knowledge that is in line with current thinking in educational research and in cognitive psychology.

SK may be loosely divided into propositional (or declarative) knowledge and procedural knowledge (Allwood, 1990; Broers, 2002; Huberty, Dresden and Bak, 1993). Propositional knowledge refers to the elementary knowledge fragments (propositions) that any student of statistics needs to know. These may be either basic or advanced. Depending on the subject and level of the course, a different collection of propositions will need to be conveyed and assessed. Propositional knowledge in itself is not enough evidence for a meaningful grasp of statistics. Statistics can only be actively used where students learn to see connections between various important concepts and principles. The perception of interrelationships between concepts, or the ability to link various propositions is taken to reflect conceptual or connected understanding (Broers, 2002; Huberty *et al*., 1993; Kelly, Sloane and Whittaker, 1997; Schau and Mattern, 1997). Procedural knowledge, lastly, refers to the ability of students to correctly apply statistical procedures. This may refer to simple computational skill or the correct execution of an analysis of variance, but it may also refer to more subtle abilities like being able to infer from a general description of a research project which research design was used and what type of statistical analysis should be deployed.

Insofar as SL, SR and ST are defined in terms of SK, they seem to lay different emphasis on the above dimensions of SK. SL is usually portrayed in terms of propositional knowledge. As has been discussed above, there is considerable dispute on what kind of propositional knowledge is required in order for someone to qualify as statistically literate, but it seems feasible to couch the objectives of SL in terms of propositions: a list of knowledge items that students of statistics need to know. Any individual researcher or statistics educator can decide for him or herself which particular propositions should be mastered in a given educational setting. These demands may vary depending on the particular target population and the level of the course. SR, on the other hand, is more concerned with conceptual understanding or the ability to perceive interrelationships between various statistical concepts and ideas. The items of the SRA are particularly suited to assess the extent to which students have been able to perceive links between important propositions. The items represent only a small sample of potential subdomains of SK, and again depending on the particular group of students and level of the course they could be supplemented by similarly structured items covering a host of different subdomains of SK. Lastly, the ST model developed by Wild and Pfannkuch (1999) describes what an expert statistician actually does when applying statistics. It is an ideal that can only be reached by repeated exposure to problem-contexts that call for the meaningful application of statistics. The tasks presented by Chance (2002), aimed at stimulating and assessing the ability to think statistically are all tasks that demand the interplay of conceptual understanding and procedural knowledge of statistics. Careful description of the particular operations (types of procedural knowledge) and corresponding conceptual understanding that we want the students to master, may help to structure our teaching in a way that will meet the desirable learning goals. These goals will focus on types of SK that reflect ST.

To summarize: SL, SR and ST are intuitively plausible concepts that help us to structure discussions on the future direction of statistics education. What SL, SR and ST have in common is that they all refer to a core of SK. On top of that, these domains for learning possess a rather ephemeral surplus value that has intuitive appeal but in practice cannot be translated into a measurable learning goal. Where concrete learning goals are concerned, we should always demarcate to which subdomain of SK the goals apply and seek to build tasks and assessment instruments that will help students to actively explore the interrelationships among the concepts and ideas of this subdomain (reasoning, resulting in the development of conceptual understanding). Additional tasks will help the student to gain the necessary procedural knowledge which, together with conceptual understanding, will help the student gradually to evolve in his mastery of statistics.

REFERENCES

Allwood, C. M. (1990). On the relation between justification of solution method and correctness of solution in statistical problem solving. *Scandinavian Journal of Psychology*, 31, 181-190.

Broers, N. J. (2002). Selection and use of propositional knowledge in statistical problem solving. *Learning and Instruction*, 12(3), 323-344.

Broers, N. J. and Imbos, T. (2005). Charting and manipulating propositions as methods to promote self-explanation in the study of statistics. *Learning and Instruction*, 15(6), 517-538.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3).

Cobb, G. (1992). Teaching statistics. In L. A. Steen (Ed.) *Heeding the Call for Change: Suggestions for Curricular Action*, MAA Notes #22, (pp. 3-43). Washington, DC: Mathematical Association of America.

delMas, R. C. (2002). Statistical literacy, reasoning and learning: A commentary. *Journal of Statistics Education*, 10(3).

Gal, I. (Ed.) (2000). *Adult Numeracy Development: Theory, Research, Practice*. Cresskill, NJ: Hampton Press.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht: Kluwer Academic Publishers.

Garfield, J. (1999). Thinking about statistical reasoning, thinking and literacy. Paper presented at First Annual Roundtable on Statistical Thinking, Reasoning and Literacy (STRL-1).

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3).

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal,* 2(1), 22-38.

Huberty, C. J., Dresden, J. and Bak, B. (1993). Relations among dimensions of statistical knowledge. *Educational and Psychological Measurement*, 53, 523-532.

Kelly, A. E., Sloane, F., and Whittaker, A. (1997). Simple approaches to assessing underlying understanding of statistical concepts. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 85-90). Amsterdam: IOS Press and the International Statistical Institute.

Moore, D. S. (1992). Teaching statistics as a respectable subject. In F. Gordon and S. Gordon (Eds.), *Statistics for the Twenty-first Century*, MAA Notes #26, (pp. 14-25). Washington, DC: Mathematical Association of America.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3).

Schau, C. and Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal and J. B. Garfield (Eds.). *The Assessment Challenge in Statistics Education*, (pp. 91 - 104). Amsterdam: IOS Press and the International Statistical Institute.

Snee, R. D. (1990). Statistical thinking and its contribution to total quality. *The American Statistician*, 44, 116-121.

Snell, L. (1999). Using *Chance* media to promote statistical literacy. Paper presented at the 1999 Joint Statistical Meeting, Dallas, TX.

Ullman, N. (1995). Statistical or quantitative thinking as a fundamental intelligence. Unpublished paper, County College of Morris, Randolph, NJ.

Wallman, K. K. (1993). Enhancing statistical literacy: enriching our society. *Journal of the American Statistical Association*, 88, 1-8.

Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal and J. B. Garfield (eds.), *The Assessment Challenge In Statistics Education*, (pp. 107-121). Amsterdam: IOS Press and the International Statistical Institute.

Wild, C. J. and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223-265.