# ASSESSING STUDENTS' UNDERSTANDING OF STATISTICS

Luc Budé
Maastricht University, The Netherlands
Luc.Bude@stat.unimaas.nl

*Statistical literacy, reasoning, and thinking may be the most prominent objectives of statistics education; they are unsatisfactorily defined and demarcated. Therefore, they are difficult to monitor, and assess. As a consequence they are impractical as educational goals. Instead, assessment could be focused on those aspects of specific statistical knowledge that are indicative for different levels of understanding. Factual knowledge directly derived from sources of information indicates a superficial level of understanding; a comprehensive, coherent knowledge structure indicates a more profound level of understanding, and the ability to transfer knowledge (the ability to flexibly engage statistical knowledge in novel tasks) indicates an expert level of understanding. This classification of hierarchically related levels of statistical understanding may produce adequate ways of measurement and assessment.*

## INTRODUCTION

Traditional approaches to teaching statistics used to emphasise skills, procedures and computations. Over the past decade there has been a shift toward stimulating a more profound level of understanding of the subject matter. For modern statistics education statistical literacy, statistical reasoning and statistical thinking are introduced as more desirable learning outcomes. Yet, there is no formal agreement regarding the definitions and distinctions of these terms (Ben-Zvi and Garfield, 2004). Some excellent attempts have been made to define and demarcate this terminology, still there exists considerable confusion. This confusion in the terminology causes major problems in the assessment of students and consequently in determining whether the educational goals are attained.

Gal (2002) has made an outstanding attempt to define and demarcate part of the terminology. He presents a comprehensive conceptualisation of statistical literacy. His definition is: the ability to interpret, critically evaluate, and communicate about statistical information. He claims that statistical literacy is more than minimal statistical knowledge; it encompasses certain skills, beliefs, attitudes and a critical perspective. However, a problem is that he does not provide a cross-reference to statistical reasoning and/or thinking. A second problem is that according to Gal statistical literacy comprises the ability to communicate and discuss over opinions and understanding of statistical information and concerns about conclusions. This implies that in his view statistical literacy is a much broader concept than in the conceptualisation of other authors (Ben-Zvi and Garfield, 2004; delMas, 2002; Rumsey, 2002; Snell, 1999). Moreover, the model he presents is more applicable to data consumers than data producers, e.g., students in enquiry contexts. This seems impractical for a concept that is postulated as an important learning outcome.

Other problems regarding the confusion and obscurity of the concerning terminology pertain to the wording of the definitions. Some definitions are circular. For example, statistical reasoning is sometimes defined as the way people reason with statistical ideas (Garfield and Chance, 2000; Ben-Zvi and Garfield, 2004). Other definitions use terms like the ability to understand, the ability to fully interpret and judge statistical results, the appreciation of the concepts, etc (Ben-Zvi and Garfield, 2004; delMas, 2002; Garfield, 2002; Rumsey, 2002; Snell, 1999, Walman, 1993; Wild and Pfannkuch, 1999). The questions that can be raised are: What is understanding? What is the ability to interpret? What is appreciation? These are all fuzzy and ill-defined concepts themselves. They do not contribute to the clarification of other obscure terms.

Furthermore, the terms statistical literacy, reasoning and thinking are used in a general way. They are in fact not related to specific statistical topics. However, proficiency in some domain is always related to certain topics. A student can have some expertise on graphs, *t*-tests, and analysis of variance; he may be ignorant about logistic regression and factor analysis. The fact that the terms statistical literacy, reasoning, and thinking are not related to specific statistical topics adds to the fuzziness and the impracticability of these terms as learning outcomes.

The fuzziness and obscurity of the terms is yet aggravated, because there does not even seem to be agreement on whether literacy, reasoning, and thinking are hierarchically related. Ben-Zvi and Garfield (2004) provide definitions that do indicate a vertical hierarchy, statistical literacy being basic and statistical thinking being the highest level. Gal (2002) does not specify how statistical reasoning and statistical thinking are situated in relation to statistical literacy. Chance (2002) presents components of statistical thinking but does not order literacy, reasoning and thinking hierarchically either. It is claimed that these terms are often used interchangeably Chance 2002; delMas, 2002). Rumsey (2002) treats statistical literacy as equivalent to reasoning and thinking and proposes to replace these terms by the hierarchical terms statistical competence and statistical citizenship.

What may have contributed to this obscurity is the fact that statistical literacy, reasoning, and thinking are treated as theoretical constructs, even though they are not (N.J. Broers, personal communication, 14 October, 2005). Broers claims that they did not originate from empirical regularities that needed theoretical explanation. Instead, they are created out of the desire to formulate more modern and meaningful educational goals for statistics education. This has lead to diverse and in a sense arbitrary conceptualisations.

As a consequence of the above mentioned problems the debate on the terminology is far from over, no general consensus has been reached yet (Watson 2002; Ottaviani, 2002; Batanero, 2002). Obviously, this has far-reaching effects on the assessment of statistical literacy, reasoning, and thinking. In the following I will discuss some approaches of statistical assessment and after that an alternative will be presented.

## ASSESSMENT OF STATISTICAL PROFICIENCY

The postulation of statistical literacy, reasoning and thinking as more desirable learning outcomes than mere skills, the correct use of procedures and computations, implies the necessity of some kind of assessment of these concepts. However, the respective definitions of these different domains of statistical proficiency are ambiguous, there is considerable overlap, and the concepts are impractical for everyday statistics education and assessment.

In spite of this, several excellent approaches to the assessment of statistical proficiency have been put forward. Nevertheless, as a consequence of the aforementioned, these approaches are inherently problematic. For example, Garfield (1998) presents an instrument for assessing statistical reasoning, but this instrument only measures a small subset of reasoning skills. It is focused on data, graphics, statistical measures, uncertainty, samples, and associations. It does not measure the application of reasoning skills on more advanced statistical techniques (Garfield and Chance, 2000). So it is not suited as an instrument to measure whether a person is able to fully interpret statistical results of these more advanced techniques.

Chance (2002) presents examples of how statistical thinking could be assessed. However, she puts that it is difficult to differentiate between statistical reasoning and thinking. Other assessment approaches seem not to be related to the postulated learning outcomes. It is suggested that assessment should be directed at students' ability to construct or interpret statistical arguments and understanding of the logic. Methods of assessment should gauge the assimilation of students' skills, knowledge, dispositions, and their ability to manage meaningful realistic questions, problems or situations (Gal and Garfield, 1997). These recommendations are not directly related to statistical literacy, reasoning or thinking. Probably this fact reflects how ineffective these concepts are as learning outcomes.

It can be concluded that an accurate method of assessment of statistical literacy, reasoning or thinking does not exist. Nevertheless, these attempts show that there is a need for a fine-grained assessment procedure. This procedure should enable differentiation (1) between topics and (2) between educational goals toward these topics. First, this means that assessment has to be focused on a specific topic, e.g., a statistical technique. As explained, proficiency in statistics is always related to a specific topic. Accordingly assessment should not be focused on general competencies, but it has to reveal what a student has understood of a specific topic. For example, it should reveal whether students have understood everything that is related to for example multiple linear regression, a *t*-test, or factor analysis, the underlying assumptions, when application is appropriate, if they are able to interpret the results, etc. In the following will be claimed that the contents of an individuals knowledge structures mirror what that individual knows about such topics.

Secondly, assessment should also demonstrate whether the educational goals toward a specific topic are met. In different educational settings different levels of understanding might be desirable. The educational goals concern the desired level of understanding. This implies an assessment procedure that can differentiate between these desired levels of understanding. In cognitive psychology some qualities and characteristics of knowledge structures are attributed to different levels of understanding. So, not only what an individual knows about some topic is mirrored by her/his knowledge structures, the level of understanding is determined by an individual's knowledge structures as well

Therefore, a fine-grained assessment procedure directly aimed at students' knowledge structures not only provides an assessment procedure that is related to specific topics, it can also reveal hierarchically related levels of understanding with respect to those topics. It is claimed that such an approach is more fruitful than to strive for assessing statistical literacy, reasoning and thinking.

## ASSESSMENT OF THE LEVEL OF UNDERSTANDING

Understanding of subject matter is a continuum. It ranges from no notion of the subject matter at all, to a complete and exhaustive insight of the material. Yet, based on cognitive psychological principles a ranking in three levels is proposed and will be discussed. This taxonomy is not only based on cognitive psychological principles, it also in line with the tendency to postulate three levels of statistical proficiency in the domain of statistics education. For example, Ben-Zvi and Garfield (2004) treat statistical literacy, reasoning and thinking as three levels; Shamos (1995) suggests so called building blocks: basic vocabulary, understanding of the science process, and understanding of the impact on society; Watson (1997) identifies three tiers: basic understanding of terminology, embedding of statistical language in a wider context, the development of a questioning attitude toward more sophisticated concepts.

The three levels of understanding that I propose are:

1. A superficial understanding based on knowledge directly derived from sources of information.
2. A more profound level of understanding based on a comprehensive and coherent knowledge structure.
3. The highest level of understanding which is indicated by the ability to transfer knowledge; i.e., the ability to flexibly engage statistical knowledge in novel tasks.

Knowledge can be conceptualised as a network of meaningful elements that are interrelated; i.e., a knowledge structure. This means that elements of knowledge are spread out in meaningful space, connected by similarities. In cognitive psychology certain features of knowledge structures are believed to indicate different levels of understanding. First, knowledge structures are believed to be restricted if they are directly derived from textbooks, lectures, etc, without further elaboration. This kind of knowledge can be correct, but limited (Kintsch, 1998), indicating a superficial level of understanding. This level of understanding is typified by isolated knowledge of definitions, formulas, algorithms, procedures, etc.

Active elaboration of course material will, secondly, result in a deeper level of understanding, mirrored by comprehensive knowledge structures that contain all the relevant concepts of the domain, as well as their relationships (Chi, Feltovich, and Glaser, 1985; Wyman and Randel, 1998). Consequently, this level of understanding is typified by the ability to combine concepts, to explain how concepts are related to each other, etc.

Thirdly, expert knowledge, i.e., the highest level of understanding, is characterised by the ability to successfully solve problems different from problems that were presented during teaching (Glaser, 1993). Solving this kind of problems asks for the transfer of knowledge. The ability to correctly transfer knowledge asks for knowledge about procedures, principles and conditions of the correct application of the learned subject matter. In other words the knowledge structures are characterised by the interrelation of procedures, principles, and conditions with all the relevant concepts of the domain.

All the three proposed levels of understanding are related to features of a person's individual, internal knowledge. This creates a problem, because internal knowledge is not directly accessible. The key problem is how to validly assess this knowledge, because this implies to validly elicit, externally represent, measure, and document this knowledge. For each of the three levels of understanding a different elicitation technique might be appropriate.

Superficial understanding, i.e., the knowledge is directly derived from sources of information, enables a person to give definitions, use algorithms in an ignorant way, and apply procedures without further knowledge. To assess this level of understanding of specific topics it suffices to ask for definitions, multiple choice questions, or asking for some basic procedural actions like filling in a formula and carrying out a computation. For example asking to compute Cronbach's α using the Spearman-Brown formula, when the mean inter-item covariance for a scale is given. In modern statistics education this level of understanding, although relevant as a basis for deeper understanding, is thought to be inadequate.

A currently desired, more profound level of understanding is conditional upon a comprehensive knowledge structure with all the relevant concepts and their relationships. A rich structure enables a person to explain phenomena and causal mechanisms, to clarify how concepts are related, and provide self contrived examples (Chi, Feltovich, and Glaser, 1985; Wyman and Randel, 1998). In the domain of statistics education this opinion is well accepted too. The ability to explain what can be inferred from data is a good indicator of whether a student understands statistics (Joliffe, 1977). For a better understanding of statistics teaching should be more directed at developing connections among concepts (Watson, 2002). For the assessment of this level of understanding open ended questions regarding specific topics are well suited. In particular, open ended questions that ask for how concepts are related to each other. For example could be asked how sample size relates to the width of confidence intervals. Answers to such questions reveal the structural aspects of the student's knowledge. The answers show how students have linked concepts, how closely related the concepts are, and how several elements of the domain can be combined (Dochy, 2001; Gijbels, Dochy, van den Bossche and Segers, 2005).

For the highest level of understanding it has to be shown that statistical knowledge can be flexibly engaged in novel tasks. For the assessment of this ability so called transfer questions are most appropriate (Bassok, 2003; Campione and Brown, 1990; Mayer, 1997). Specifically, far transfer questions, i.e., open ended questions that specifically ask for the application of particular knowledge in a different situation, different but similar problems, or a new context. Essential is that neither the question nor the answer were presented during teaching. It could be asked for example how given data, from a given study should be analysed. Students have to apply knowledge about designs, variable types, research questions, statistical techniques, etc., to a new presented situation. The ability to answer this kind of questions is dependent on the notion of how domain knowledge is connected to conditions of appropriate applications.

The assessment of superficial understanding, i.e., evaluating definitions, multiple choice questions, and correct computations is straightforward. The assessment of higher levels of understanding might be more difficult. First, it is essential that the appropriate questions are asked. The structure of knowledge is only revealed when students are asked to explain, relate, and/or combine concepts. Formulating transfer questions has to be done conscientiously too. Knowledge of other domains may contaminate the results, if for example the new context is shifted too far away from the original context. Secondly, the scoring is not effortless, because objectively scoring open ended questions is believed to be problematic. Judging these questions is believed to be both subjective and laborious. Using a detailed answer key may be a practical and reliable method (Budé, 2006).

## CONCLUSION

Statistical literacy, reasoning, and thinking are widely accepted in statistics education as learning outcomes. Because the majority of the persons involved in statistics education recognises that these terms are ambiguous, attempts have been made to demarcate and define these terms. However, there is no consensus yet. So far definitions are fuzzy and overlap. The terms are used in a global way; there is no coupling to specific topics. All these aspects make the terms statistical literacy, reasoning and thinking unfit for educational practices, as educational goals, and for assessment purposes. It is claimed that focussing directly at relevant aspects of students' knowledge structures will result in practical applicable assessment possibilities. Assessment should be focussed on students' knowledge of specific topics and try to gauge their understanding of the subject matter. Characteristics of students' knowledge structures can be associated with three levels of understanding. Isolated knowledge analogous to information as it was presented during teaching is typical for superficial understanding. Rich knowledge with numerous connections between the concepts indicates a deeper level of understanding. The deepest level of understanding is characterised by the ability to transfer knowledge. Asking for a definition or a computation is a suitable way of assessment for superficial understanding. For deeper levels of understanding explanation and transfer questions are most suited. Scoring answers to these kinds of questions can reliably be done with an answer key.

REFERENCES

Bassok, M. (2003). Analogical transfer in problem solving. In J. E. Davidson and R. J. Sternberg (Eds.), *The Psychology of Problem Solving*. Cambridge: Cambridge University Press.

Ben-Zvi, D. and Garfield, J. (2004). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking.* Dordrecht: Kluwer Academic Publishers.

Budé, L. M. (2006). A procedure for measuring understanding. Manuscript submitted for publication.

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

Campione, J. C. and Brown, A. L. (1990). Guided learning and transfer: implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, and M. G. Shafto (Eds). *Diagnostic Monitoring of Skill and Knowledge Acquisition.* Hillsdale: Lawrence Erlbaum Associates.

Chi, M. T., Feltovitch, P. J. and Glaser, R. (1985). Categorisation and representation of physics problems by experts and novices. *Cognitive Science,* 5, 121-152.

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

Dochy, F. (2001). A new assessment era: different needs, new challenges. *Research Dialogue in Learning and Instruction,* 2, 11-20.

Gal, I. (2002). Adults' Statistical literacy: Meanings, components, responsibilities. *International Statistical Review,* 70, 1-51.

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.

Gijbels, D., Dochy, F., van den Bossche, P., and Segers, M. (2005). Effects of problem-based learning: a meta-analysis from the angle of assessment. *Review of Educational Research,* 75, 27-61.

Glaser, R. (1993). *Advances in Instructional Psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jonassen, D. H., Beissner, K., and Yacci, M. (1993). *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition.* Cambridge: Cambridge University Press.

Mayer, R. E. (1997). *Thinking, Problem Solving, Cognition*. New York: W. H. Freeman and Company.

Olson, J. and Biolsi, K., J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson and J. Smith (Eds.). *Toward a General Theory of Expertise.* Cambridge, UK: Cambridge University Press.

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

Shamos, M. H. (1995). *The Myth of Scientific Literacy*. New Brunswick, NJ: Rutgers University Press.

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

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*. 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.

Wyman, B. G. and Randel, J. M. (1998). The relation of knowledge organization to performance of a complex cognitive task. *Applied Cognitive Psychology*, 12, 251-264.