LEARNING STATISTICS BY MANIPULATING PROPOSITIONS

Nick J Broers
Maastricht University
The Netherlands

Following a course in elementary statistics, students are able to demonstrate a basic knowledge of statistical concepts and ideas, but often fail to apply this knowledge to concrete problems. From research in cognitive psychology, we know that the organization of knowledge starts with the mental storage of initially isolated concepts and simple principles. A certain amount of conceptual understanding is reached when the student succeeds in forming relationships between these knowledge elements. The task faced by any teacher in statistics, is to enable the student to form such integrated knowledge networks. Research has shown that the formation of such networks is stimulated when students, confronted with a statistical problem that requires the application of their basic knowledge, actively try to explain the solution of the problem to themselves. This paper discusses a didactic method that seeks to stimulate such self explanatory activity in students.

INTRODUCTION

Teaching statistics to students with a non-mathematical background can be a demanding task. No matter how much effort we invest in making our teaching accessible and lively, in the end the student still has to learn an array of highly abstract concepts and to comprehend how these concepts interrelate. The fact that students often show avoidence behavior when confronted with statistics and frequently display a lack of motivation (see Gal & Ginsburg, 1994) makes this goal hard to achieve.

Indeed, it has frequently been found that many students have succeeded in storing a number of different concepts and principles of statistics, but that they do not know how these elementary knowledge fragments relate to each other or to problems to which they should be applied. In keeping with traditional terminology of cognitive psychology, I will refer to these elementary knowledge fragments as propositions, and to the establishment of a network of propositions as conceptual understanding (see e.g. Kintsch, 1998). Thus, the aforementioned students demonstrate a sort of propositional knowledge, but they lack conceptual or connected understanding of the statistics material (Broers, in press; Huberty, Dresden & Byung-Gee, 1993; Schau & Mattern, 1997). The problem is that difficulty with abstraction, lack of motivation and a sense of apprehension all combine to produce the end result of limited propositional knowledge and a lack of conceptual understanding in many students. As has been shown in studies on the acquisition of knowledge on physics (Chi, Bassok, Lewis, Reimann & Glaser, 1989) and biological knowledge (Chi, DeLeeuw, Chiu & LaVancher, 1994), the key to expanding limited propositional knowledge to conceptual understanding is self explanation by the student. I have been working on a didactic method that aims to stimulate the student to do just that.

OUTLINE OF THE METHOD

What many of us do when discussing a statistical topic in class, is to introduce a problem that presents a context for the exposition of the statistics. Whilst listening to a lecture or reading a text, the student, in order to comprehend what is being taught, has to 1) try to identify relevant knowledge elements, to be mentally stored as propositions, 2) attempt to connect these propositions into a coherent knowledge framework such that 3) upon presentation of new but related problems, he will know how to proceed in order to find a solution. In practice, many students do not get beyond the first of these steps, and even the storing of elementary propositions is often accomplished in an incomplete and partially distorted way (resulting in misconceptions). The method I have been working on tries to assist students in successfully taking the first of these two steps.

To help students identify the relevant propositions, we must ourselves be aware of what these propositions ought to be. Of course, this depends on the topics that we are teaching in a particular course. So far, I have only been actively working with this method in a course on descriptive statistics, amongst other topics covering a discussion on linear transformations, with centering and standardizing as specific examples. The theme of linear transformation was discussed in a single class, and what I did was carefully go through my notes to see what specific propositions I was trying to get across. Such a collection is of course limited, with a great amount of time being devoted to the explanation of these propositions, often with help of some concrete example. Table 1 shows some examples of propositions that I identified.

Table 1
Examples of propositions

Proposition	
1	A linear transformation is a transformation of the type X' = bX + a.
2	When we apply a linear transformation to the scale values of a variable X, the effect of multiplying all scale values with a factor b is that all values, as well as all distances between values become b times as large.
3	After a linear transformation the shape of the original distribution remains unaltered.
4	The value of the mean after linear transformation of X can be found by inserting the original mean into the transformation formula: $\overline{X}' = b\overline{X} + a$
5	Centering is a transformation whereby we express the score of an individual as a deviation from the mean: $X' = X - \overline{X}$
6	Centering is a linear transformation of scores with b = 1 and a = - $\overline{X}$
7	Standardizing is a transformation whereby we express the score of an individual as a deviation from the mean, relative to the standard deviation of the scores: $Z = \frac{X - \overline{X}}{s_x}$
8	Standardizing is a linear transformation of scores with b = $\frac{1}{s_x}$ and a = $-\frac{\overline{X}}{s_x}$

There were of course various other propositions covering the topic of linear transformations. What the propositions listed in Table 1 exemplify, is that although they are fairly detailed and specific, they also leave a fair amount of more basic propositions implicit. For instance, in proposition 5 we note that mention is made of the mean and its symbol $\overline{X}$; in proposition 7 mention is made of the standard deviation and its symbol s_x. Additional propositions could be listed giving details on the mean and standard deviation. In fact, propositions on the standard deviation may contain references to the square root of the variance, to the idea of variation, to the sum of squares, to the mean, etc. These in their turn contain implicit propositions on squares, square roots, possibly on degrees of freedom, etc. The list is endless and could be extended to material that is non-specific to statistics and goes back to primary school. Of course, that is not the way we teach a course on statistics (or indeed on any topic). When discussing standardization, knowledge of the mean and the standard deviation are assumed to be already familiar to the student, and thus needing no further attention. A student taking our course is assumed to possess relevant prior knowledge and therefore will himself not make use of any of these more elementary propositions. The list of propositions that we compile should be a list of all those elementary statements that are truly novel to the student, and that he needs to encode or mentally store for the first time.

After I had completed my list of basic propositions, I started to convert each of these into questions for the students. Questions like: *The distribution of X has a certain shape. How will this shape change after linear transformation of X?* (proposition 3) and *How does the value of the mean change after linear transformation of X?* (proposition 4). The student is therefore not handed the relevant propositions on a plate. Instead he is urged to study all the course material – lecture notes as well as books or readers – and to search for answers to these questions. Without these questions the students would have to scan all the material and to decide for themselves what the relevant knowledge elements are. Especially for weaker students this may be a disheartening task, as the confrontation with the elaborate body of unknown and often abstract material soon make it hard for them to see the wood for the trees. By structuring the material, we help them making the trees clear again.

After they have identified the relevant propositions, they can try to memorize these, but obviously that would not help in achieving the second step of the learning process: establishing cross links between the knowledge elements. To help them construct those links and thus perceiving relationships among the important concepts, the second phase of the didactic method consists of presenting the students with short problems that require the interrelated use of propositions for their solution. This took the following form. I presented all students with a number of small texts containing statistical information, each text being followed by four true/false questions. The true/false questions were accompanied by those propositions that should be used in order to derive the answer. The student was instructed to construct an argument on the basis of the provided propositions , that would show the question to be either true or false. An example of such a text is the following:

> The distribution of a variable X is shown to be positively skewed. We transform the X-scale into a new scale X'. The distribution of the X' scores has a symmetric shape. The mean of X' is zero and the standard deviation of X' is twice as large as the original standard deviation of X. Show the following questions to be either true or false, using the accompanying propositions.

Two of the four true/false questions that followed this text were:

a) The value of the mean of X' shows us that the transformation used was not centering (*use proposition 5: Centering is a transformation whereby we express the score of an individual as a deviation from the mean:* $X' = X - \overline{X}$ *and proposition 4: The value of the mean after linear transformation of X can be found by inserting the original mean into the transformation formula:* $\overline{X}' = b\overline{X} + a$)

b) The X' scores are not standardized or centered (*use proposition 3: After a linear transformation the shape of the original distribution remains unaltered, proposition 6: Centering is a linear transformation of scores with b = 1 and a =* $-\overline{X}$ *and proposition 8: Standardizing is a linear transformation of scores with b =* $\frac{1}{s_x}$ *and a =* $-\frac{\overline{X}}{s_x}$)

By focussing on the relevant information in the text of the problem and combining this with the provided propositions, the above two true/false questions can be shown fairly simply to be respectively false and true. Of course, more complicated questions can be developed that require many more than just two or three propositions. The fact that we have analyzed our course material into its constituent propositions, makes it possible for us to consciously design a number of questions of ever increasing difficulty.

I have so far tried out this method in a few pilot studies and in a single major experiment. Does it work? The preliminary results appear promising. What students appreciate especially is the help that the method provides in seeking out the relevant propositions. It gives them a feeling of control and oversight, thus making the statistics course seem less impenetrable and therefore increasing the self confidence of the students.

As to the effect of the method on the development of conceptual understanding, there is some indication of a slightly improved performance on statistical aptitude tests. Students who had been trained using the method, scored better than controls on a test assessing purely propositional knowledge (or “facts”), as well as on a test assessing conceptual understanding of the material. The differences in performance were not significant however, but the strength of the experiment was hampered by the use of small groups (ranging from 10 to 13). Personal interviews indicated that students appreciated the didactic approach as a means to structure and elaborate their knowledge of statistics.

REFERENCES

Broers, N.J. (in press). Selection and use of propositional knowledge in statistical problem solving. *Learning and Instruction.*

Chi, M., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, *13,* 145-182.

Chi, M., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education* [Online], 2 (2).

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

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge: Cambridge University Press.

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I.Gal and J.B. Garfield (Eds.), *The assessment challenge in statistics education.* Amsterdam: IOS Press.