Spontaneous concept maps aiding the understanding of scientific concepts

Virpi Slotte and Kirsti Lonka, Learning Centre, University of Helsinki, Finland

This study evaluated concept maps spontaneously constructed by applicants (N = 502) in a medical school entrance examination. In all, 36 maps were produced. Concept maps were evaluated for content of relevant terms and for the number of interrelationships indicated. The aim was to determine whether including relevant ideas on a concept map is related to the learning of those ideas. Because concept maps are effective tools for making the structure of knowledge explicit, it was hypothesized that the quality and content of spontaneously made maps would be related to improvement in the comprehension of text material.

Understanding was assessed in terms of success in essay-type tasks designed to measure the ability to define, explain, and apply statistical knowledge. The results indicated that merely including the relevant concepts in a map has little effect on the comprehension of those concepts, whereas the extent and complexity of concept maps plays a powerful role in the understanding of scientific texts.

Introduction

There is evidence suggesting that concept mapping provides a theoretically powerful and psychometrically sound tool for assessing conceptual change in experimental and classroom settings (Markham et al. 1994). However, not much is known about spontaneously made concept maps, that is, maps constructed by students who use this graphic metacognitive tool without its being experimenter-imposed. It is possible that the instructions given by researchers limit or interfere with students' customary approach to learning, or that the relationship between mapping and improved comprehension is dependent on the instructional time needed to teach students to construct maps effectively. Again, it is not always known in experimental situations how seriously students try to understand the text. Therefore, we need to examine spontaneous drawing of concept maps in real, ecologically valid, learning situations (Mayer 1992).

The present study looks at the concept maps that were spontaneously constructed in a highly demanding and motivating situation, namely, in an examination taken for admission to medical school. The purpose is to determine how qualitatively different types of maps are related to understanding scientific concepts. Moreover, this study aims at gathering further data on concept maps used in a natural learning setting, a situation in which the effects of spontaneous maps have not yet been reported.
Learning constructively

Basic to science learning is the development of a deep understanding of the central science concepts necessary to form an extensive and well-organized knowledge base. Since scientific knowledge is fairly nonlinear and web-like, concepts should be learned as organized networks of related information, not as lists of facts (Glynn and Muth 1994, Fellows 1994). Therefore, students need to challenge the science text they read by struggling with it and trying to make sense of the subject matter. They do this by selecting and organizing relevant information and making links between concepts. In addition, the self-regulated learner must appropriately control his or her learning processes by integrating new knowledge with what he or she already knows. Thus, cognitive learning theory emphasizes the constructive nature of reading and comprehension (Glaser 1991, Mayer 1992).

One tool for fostering the construction of conceptual relations is concept mapping. A concept map is a two-dimensional representation of information which illustrates the connectedness between and among individual concepts (Novak 1993). Making a map requires a learner to become actively involved in identifying the central idea and relating them to each other in a meaningful way (Ausbuchel 1968, Heinze-Fry and Novak 1990). This involves the additional processing of material by learner-generated activity (O’Donnell 1994) resulting in the compact, efficient and easily accessible storage of information in memory (Anderson 1985, Stensvold and Wilson 1990).

The ability to make links between concepts and ideas may indicate the ability to synthesize related statements about how a certain piece of the world looks or works. Also, the links may show a new interpretation of old ideas and some degree of creative thinking. Therefore, mapping can lead not only to both a quantitative and a qualitative increase in the learner’s knowledge structure, but also to misconceptions. Very powerful links might relate previously remote chapters or even subject areas, whereas rote learning merely involves a series of propositions that are memorized, but not related to each other (Ausbuchel 1968).

Because of the relationships between concepts, maps reflect the psychological structure of knowledge rather than a linear text (Wandersee 1990). Examples of study techniques that reflect the logical structure of text are outlines, notes and underlining. However, students often use underlining and outlining to familiarize themselves with the scientific concepts while reading the text. Therefore, the first version of maps may also contain lists of facts, whereas the construction of conceptual relations that form the basis of real scientific expertise has probably undergone numerous revisions. In this study, spontaneous concept maps are defined as naturally made maps containing all kinds of graphic representations consisting of a minimum of three links or relationships between the concepts (Lonka et al. 1994).

Concept maps and mental representations

Designed to parallel human cognitive structure, concept maps seem quite appropriate for the holistic representation of scientific knowledge (van Dijk and Kintsch 1983, Wandersee 1990) distinguish three qualitatively different forms of mental representation: (1) surface memory for actual words and phrases, (2) a textbase, in which a coherent representation of the text is formed, and (3) a situation model, in which the text content is integrated into the comprehender’s knowledge system.
The textbase reflects the coherent relations between the propositions in the text and their organization, whereas the situation model is a mental representation of the situation described by the text. The textbase and the situation model are not independent of each other, but each has its own characteristics.

Mannes and Kintsch (1987) provided empirical evidence suggesting that forming a situation model results in superior performance in creative problem-solving tests that require deeper understanding of the material. Previous research also indicates that spontaneous concept mapping is useful when learning is being assessed on the basis of essay-writing requiring the formation of a situation model. These findings were replicated in two separate studies on a so-called learning-from-text test, which is one part of an entrance examination for medical school in Finland (Lonka et al. 1994, Lahtinen et al. 1997). In neither of these studies, however, were the quality and content of naturally made maps taken into account.

Nevertheless, a number of research studies on imposed maps have shown that more complex concept maps help in organizing and understanding new subject matter better than less complex maps (Heinze-Fry and Novak 1990, Starr and Krajcik 1990, Pankratius 1990, Willerman and MacHarg 1991, Mahler et al. 1991, Okebukola 1992, Roth and Roychoudhury 1993, Horton et al. 1993). Good concept mappers were also shown to be better problem-solvers in science (Okebukola 1992).

Stensvold and Wilson (1990) found that while the number of words students wrote in a map was not related to their test performance, the number of valid links on it predicted comprehension test performance. A study by Wallace and Mintzes (1990) showed that students who constructed the most complicated maps were able to generate a substantially greater number of scientifically acceptable propositions concerning marine life zones than those who drew the simplest maps. Correspondingly, the work by Markhan et al. (1994) indicated that concept maps of biology majors were structurally more complex than those of non-majors.

Taken together, the concept-mapping effect has ample empirical support from previous research in science understanding. The mapping effect is based on the structural complexity of concept maps, reflecting the mental representation constructed from the main ideas in the text. In the present study, we wanted to know whether this effect would also occur with spontaneously made maps in a new subject domain in statistics.

**Predictions**

The threefold purpose was to determine (1) what kind of concept maps participants spontaneously construct, (2) whether the most essential concepts and propositions in participants' concept maps were also found in essay-type tasks, and (3) whether the accuracy of the maps reflects an understanding of statistical probability concepts presented in a complex text. Generally, it was hypothesized that the accuracy of spontaneously made concept maps would be related to understanding scientific concepts. This hypothesis was operationalized into the prediction that the number of appropriate concepts and their relations in a concept map would be related to improved understanding of text material.

Although most rapidly-made maps are incomplete in many ways, the accuracy of any map is highly dependent upon the quality and quantity of the data the map maker collects about the reality to be mapped (Wandersee 1990). Here, the accu-
racy of concept maps is measured by the extent (number of appropriate words and examples) and the complexity (number of valid relationships) of the mapping. Understanding is assessed on the basis of success in essay-type tasks measuring different aspects of scientific text comprehension. Theoretically, the complexity of concept maps made by students should enhance science understanding (van Dijk and Kintsch 1983, Novak 1993).

Method

Participants

The participants were all the applicants who spontaneously constructed concept maps while reading text during an entrance examination to a medical school. This type of examination is taken as part of the admission process for a 6-year study program combining medical school and graduate studies. Of 502 applicants, 36 produced concept maps.

Most of the applicants were very good students, 75% having a top grade (Laudatur) in their high school diploma. The criteria for admission to the University of Helsinki Faculty of Medicine include in addition to the applicants' prior school performance (overall matriculation examination grades), the results of three scientific knowledge tests (physics, chemistry and biology) and the so-called learning-from-text test to assess reading-comprehension and essay-writing skills.

Materials

The material of the two-hour learning-from-text test consisted of two texts called 'Risk Theories of a Single-case' and 'Finnish Health Risks' which were 5 and 12 single-spaced pages long respectively. The first was a university-level text including theoretical information about the interpretation of probability theories and risk theories, whereas the second one was a popular description of health hazards. Again, because only eight applicants made concept maps of both texts, all the analyses were restricted to the maps and essay-type answers related to the first text which was intellectually more demanding.

The text about risk theories was written for the purpose of the entrance examination and for later use as research material. The choice of text material was based on the unfamiliarity of the domain to as many applicants as possible. The text was constructed from information gathered from several sources and had four headings corresponding to the four sections in it. No subheadings appeared in the text.

The first part of the text contained the definition and examples of statistical risk; the second part emphasized the difference between probability functions that measure relative frequencies and those that measure causal propensities being derived for individual events. The inferences based on propensity theory and frequency theory were also characterized by the distinction between 'exposure to risk' and 'risk of exposure'. The concept 'exposure to risk' describes risk that is independent of general and concrete situations, whereas the extent of the 'risk of exposure' varies all the time according to a person's actions, circumstances and situation.
The third part of the text considered risk judgements in theory and in everyday reasoning. It included the formula for Bayes' theorem, according to which one should take into account base rates when making judgements about probability in a specific situation (Gardner 1985: 374). The apparent failure of the ordinary person to recognize the importance of the base rate was illustrated by Kahneman and Tversky's (1973) 'cab example'. In their experiment, subjects were told that blue and green cabs operate in a certain town in the ratio of 85 to 15, respectively. A witness identifies a cab in a crash as green, and the court is told that in the relevant light conditions he can distinguish blue cabs from green ones in 80% of cases. The median probability that the cab involved in the accident was blue was estimated as 0.2, and investigators claimed that this shows the prevalence of serious error, because it implies a failure to take the base rate (that is, prior probabilities) into account. Accordingly, despite the fact that the world operates according to Bayes' theorem, it does not correspond to the probability judgement in the everyday reasoning of the ordinary person.

The last section of the text concerned risk factors in expert judgements of single-case probability. The statistical probability concerning the whole system, i.e. diagnostic success in a hospital, was opposed to that for a concrete particular patient. The patient is concerned with success in his or her own particular case, not with stochastic success for the system. He or she therefore needs to evaluate a propensity-type probability, not a frequency-type one, and standard statistical methods would then be inappropriate.

Procedure

The participants were given their own copy of the texts and were told that they would have one hour to read them silently. The limited reading time was set to prevent applicants from learning the texts by heart and to encourage them to concentrate on the essentials. They were also informed that they could mark these copies or use an attached blank sheet of paper in any way they desired. The reading instructions, which were typed on a separate sheet of paper, indicated that essay-type answers would be scored for both content and clear presentation. Before the questions were given, all the text material and the concept maps were collected and not returned to the participants. The answers to each task had to be written within a given space on the answer sheet. The applicants were given another hour to complete the tasks.

In the first task, the participants had to define the essential points of frequency theory and propensity theory, called hereafter Definition of Frequency Theory and Definition of Propensity Theory. This task was meant to measure both recall and comprehension of these theories. The next task, the only one involving the information described in both texts (i.e. compare the accident risks of persons A and B described in the second text according to frequency theory and propensity theory), was dropped from the analysis. In the Explanation of Bayes' Theorem, the participants had to explain in their own words to a 50-year-old, poorly educated, person what Bayes' theorem means. This task required clear presentation and full comprehension of the theorem described in the text. The Application of Risk Theories involved analysing the risk theories or interpretations that may lie behind a physician's inferences when he or she makes conclusions about a disease or its prognosis. This task was meant to measure the ability to synthesize the content and go
'beyond' the text by making inferences as to how different theories could be related to the physician's work.

Analyses

Spontaneously made concept maps were scored by counting the number of valid concept words and examples, and the number of relationships between the words or concepts on each student's map. The number of concepts and relationships were taken as indications of the extent and the complexity of the maps. The sum of these terms is thought to measure the accuracy of mapping correct propositions, in other words the quality of spontaneously constructed concept maps.

One point was awarded for each appropriate relationship, concept word and example. Examples, which include drawings on a concept map, represent specificity of knowledge (Markham et al. 1994). If the same concept words were written many times, each concept was counted only once, whereas each new link showing a correct relationship between two concepts was taken into account. One point was given for any kind of valid relationship (e.g. linkage, branching, or arrow) illustrating the connectedness between the concepts that make up a proposition (Novak and Gowin 1984). However, points were not given for dashes or any other symbols in front of the listings.

The concept maps in this study displayed very few hierarchies and little cross-linking between groups of concepts. Moreover, because of the spontaneous nature of the mapping, the representations were not always unambiguous. For the most part, this was assumed to be due to the lack of time or opportunity given to participants to revise or make changes in their maps. Hence, no distinctions were made among linkages, hierarchies and crosslinks. All points were added together to obtain the final mapping score. The use of a cumulative concept mapping score is recommended by Novak and Gowin (1984).

The essay-type answers in the entrance examination were analysed using a detailed *a priori* scoring system which was applied to all answers independently by two official raters \((N = 502)\). The content scores for both the definition tasks varied from 0 to 4, and for the explanation and application tasks from 0 to 6. Three additional points were given for each task on the basis of the quality of the argument, in other words, whether the answer was understandable, logical and coherent.

The concept maps were scored separately from the responses to the essay-type tasks by two raters. Any differences in scoring were settled by discussion between the raters. The reliability rating of the final concept mapping scores was 0.93.

According to the final mapping scores, the 36 maps were divided into three percentile groups. This resulted in 12 in the below-average group (BA), 12 in the average group (A), and 12 in the above-average (AA) group. There were no significant differences between the mapping groups BA, A and AA in terms of overall matriculation examination results \((F(33, 2) \approx, p = 0.75)\). Performance in the BA, A, and AA groups in terms of the accuracy of the maps and the standard of the four essay-type tasks was compared by one-way analyses of variance.

Chi-square and Fisher's exact tests were carried out to see to what extent the relevant ideas in a concept map are related to recalling, comprehending and applying them. Fisher's test was favoured because in many cases the distributions were very biased due to the natural settings, that is, the number of the relevant ideas in
the maps could not be controlled. The test was applied whenever a cell with an expected frequency of under five appeared.

Results

What kind of concept maps were spontaneously constructed?

The results in general show that spontaneously made concept maps varied substantially in their extent and complexity. Therefore, the maps offer a preliminary insight into some of the characteristics that are frequently seen in students with different levels of conceptual understanding. The first one, which is a below-average example (figure 1), shows a minimal number of concepts, examples and linkages. 'Health risk' and 'accident risk' are considered examples of statistical risk. The definition of the 'probability of statistical risk' that is based on exposure is not clearly formed in this map. Concept words are written parallel to one another, without any connective signs. The words 'duration', 'frequency' and 'performance' are mentioned in the text as characteristic of activity in relation to which exposure to risk can be defined. Yet none of the probability theories based on statistical risk, that is Frequency theory, Propensity theory and Bayes' theorem, is mentioned. As a result, the final concept mapping score for this map was 14 (8 concept words or examples and 6 relationships).

The second concept map (figure 2) exemplifies the average mapping group. It contains more extensive textual information than the maps of the BA group, referring more often to a certain part of the text. Unrelated concepts are presented here, too, in addition to the ideas central to the risk factors concerning a single case (e.g., 'Karl Popper', 'quantum mechanics'). Organizational links between concepts are still somewhat lacking, resulting in a final concept mapping score of 24.

In the third map, which illustrates the performance of above-average mappers (figure 3), the incidence of concept words, examples and branching is higher. Above-average mappers generally produced more concepts and relationships related to the knowledge domain of statistical risk. The concept of probability is successively linked to frequency and propensity theories in this map. However, the descriptions of these theories are repeated in another direction, and only listed one underneath the other.

![Figure 1. An example from the below-average concept map.](image)
**Figure 2.** An example from the average concept map.

**Figure 3.** An example from the above-average concept map.
The notions of 'blue 85' and 'green 15' refer to Kahneman and Tversky's (1973) judgement errors of probability (see Materials). However, in the third map there are no linkages between the cab example and Bayes' theorem or the apparent failure of the participant to recognize the importance of the base rate. Bayes' theorem is mentioned, but no further differentiation is made. Only one point was awarded for the words blue 85 and green 15 because they represent the same example from the text. The numbers of concept words and examples depicted in this map is 29, and the number of valid relationship is 7.

The fourth map (figure 4) shows a tangible example of an above-average mapper containing a maximum number of concept words and examples (35) and relationships (19), resulting in a final concept mapping score of 54. Some differentiation can be seen in the interpretation of probability that is further divided into frequency interpretation based on statistical risks and propensity theory based on a single case. Furthermore, an unconnected list compares the differences between these theories. Prior experiences and correspondence with everyday reasoning are also listed under Bayes' theorem. In general, the above-average mappers used more propositions that were more highly connected describing conceptual understandings of the various components of statistical risks in their maps.

Overall, the majority of the scientific concepts and propositions in the concept maps spontaneously employed by all the mapping groups were closely related to the textual content. Applicants included only a few of their own examples and inferences in their maps. On the other hand, lists of more than two words written in the form of linear text were generally found in the maps of all groups, that is on 19 of the 36 maps (53%). This kind of listing closely resembles conventional note-taking, where concepts are typically copied in a linear fashion without integrating
the major kinds of the material (e.g. Kiewra et al. 1991, 1995, Robinson and Kiewra 1995).

The effect of mapping scientific concepts

A Chi-square test and other tests of independence were carried out to determine whether the relevant concepts and propositions included in the maps were also found in the essay-type answers. The tests revealed no significant association between the mapping effect and defining the Frequency theory; in other words, the participants recalled the relevant concepts related to Frequency theory regardless of what concepts they had included in their maps. In the Definition of Propensity Theory part, the only statistically significant difference was found among those mapping the concept of 'risk of exposure'; 68% (f = 13) of those including this concept in their concept maps mentioned it in the essay answers, whereas 71% (f = 12) of participants who did not map it failed to write about it ($\chi^2(1, N = 36) = 5.46, p = 0.02$).

Although almost half the participants included Bayes' theorem in their concept maps, only 4 of them also took the example illustrating that theorem. All of them succeeded in constructing information about Bayes' example, but the majority (72%) of the participants who did not map the example nevertheless mentioned it in their essay answers. These analyses thus did not produce significant differences, nor did any of the other analyses of this task.

In the Application of Risk Theories task, 74% (f = 25) of the applicants who presented the concept of 'Frequency theory' in their concept maps succeeded in analysing that theory in the task requiring the ability to apply information to a new situation, whereas both participants (f = 2) who omitted the concept failed to apply it in their essay answers. These analyses did not reach statistical significance (Fisher's exact test, $p = 0.09$). Consequently, none of these (f = 20) who did not include 'Bayes' theorem' in their concept maps was able to apply it in their essay answers (Fisher's exact test, $p = 0.08$). Of the 16 participants who mapped 'Bayes' theorem', only 3 succeeded in writing about it. This may be due to the difficulty of this theorem. No statistically significant difference was found among those who presented the concept of 'propensity theory'.

In summary, mapping relevant concepts had little effect on how these concepts were defined and explained. Moreover, the mere appearance of the central concepts in the maps was of only limited advantage in inferential reasoning and the application of scientific concepts, whereas omitting the relevant concepts and examples from the maps was related to the inability to apply them.

The accuracy of concept maps

The hypothesis stated that the accuracy of concept maps would be related to improved comprehension of the content matter. Since the concept maps covered all the text material, our interest focused on the impact of the quality of spontaneously made maps on success in the whole test, which consisted of various tasks. In the context of the total content scores for the essays, the above-average mapping group obtained a slightly higher mean total score (Mean = 12.92, SD = 2.99) than
Table 1. Means and SD of the concept mapping groups in the essay-type tasks.

<table>
<thead>
<tr>
<th>Concept mapping groups</th>
<th>Essay-type tasks</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Definition of</td>
<td>Definition of</td>
</tr>
<tr>
<td></td>
<td>frequency theory</td>
<td>propensity theory</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Below average</td>
<td>12</td>
<td>1.83</td>
</tr>
<tr>
<td>Average</td>
<td>12</td>
<td>2.83</td>
</tr>
<tr>
<td>Above average</td>
<td>12</td>
<td>3.08</td>
</tr>
<tr>
<td>All</td>
<td>36</td>
<td>2.50</td>
</tr>
</tbody>
</table>

Note: The total possible score for both Definition tasks was 4, for the Explanation and Application tasks 6.
the average group (Mean = 12.75, SD = 2.42), whereas the performance of the below-average group was the worst (Mean = 9.75, SD = 3.08). According to a one-way analysis of variance, these differences were statistically significant ($F(33, 2) = 4.70, p = 0.02$). Scheffé's post hoc comparison revealed that the source of the significant difference for the mapping group main effect was in favour of the above-average mappers. The average mappers also scored higher than the below-average mappers on total content ($p < 0.05$).

In order to get a wider perspective on the results, the total essay scores on the concept mapping groups in each task was further examined. The results for tasks 1–4 are presented in table 1.

Table 1 shows that the participants who made the most extensive and complex maps in the Definition of Frequency Theory part scored higher than those who constructed less accurate maps. The differences between these three groups were statistically significant ($F(33, 2) = 7.29, p = 0.002$). Interestingly, however, the average concept mappers produced the highest scores in the Definition of Propensity Theory and in the Explanation of Bayes' Theorem tasks, and the participants who made the least accurate concept maps obtained the lowest scores. Nevertheless, only the results of the analyses of variance for the Definition of Propensity Theory were significant ($F(33, 2) = 3.59, p = 0.04$). In the Application of Risks Theories task, the participants who made above-average maps scored higher than those who constructed either average or below-average maps, but the effect for this task was not significant.

*Is concept mapping advantageous in gaining admission to medical school?*

No differences in prior school performance were found between participants who constructed concept maps and those using other strategies or no strategy at all while reading scientific texts ($F(501, 1) = 1.03, p = 0.31$). Nevertheless, the proportion of applicants who were accepted was considerably higher among those who in the learning-from-text test spontaneously drew concept maps (36%) than among those who took verbatim notes (22%), or only underlined their text papers (20%). Of those who did not produce any physical marks or artefacts, only 14% passed the entrance examination. Yet, 34% of applicants who took summary notes in their own words were accepted on the programme. The differences between the strategy groups were found to be statistically significant ($\chi^2(4, N = 502) = 13.78, p < 0.01$) when all tests were taken into account (learning-from-text test, applicants’ prior school performance, three scientific tests in physics, chemistry and biology).

**Discussion**

A Finnish university examination offered a well-standardized and yet realistic setting for carrying out a study of spontaneous concept maps, the aim of which was to analyse the qualitative and quantitative data from the maps produced by applicants to the Faculty of Medicine at the University of Helsinki. Because of the highly motivating and demanding learning situation, the participants may be assumed to have done their best in trying to understand the text. Furthermore, it was assumed that those who spontaneously made concept maps considered this
strategy the most effective in understanding science. The observation that all maps were very much to the point also support the participants' motivation.

The results indicate that students who constructed a concept map achieved an academic gain over those who did not use any graphic tool while reading the text on statistics. Moreover, the extent and complexity of spontaneously made maps were related to an improvement in science understanding in the whole test. Not only are these results consistent with existing concept mapping literature, they also support the view that maps that are promoted and are constructed without immediate prior instruction are effective. The findings are also encouraging, because they indicate that spontaneous mapping improves students' performance in applied tasks.

*Spontaneously constructed concept maps*

The natural setting meant that there was substantial variation in the quality and content of the concept maps. Most spontaneous maps were not very sophisticated in terms of hierarchies and crosslinks. The reason for this may in part lie in the lack of prior systematic guidance in mapping technique, or in the tendency not to build external links between concepts but rather to learn material as discrete 'chunks' (Novak et al. 1983). On the other hand, the shortage of crosslinks in the maps may also have been caused by the limited reading time that was imposed to steer applicants' concentration towards the essential. It is possible, for instance, that those who are low in reading skills did not have enough time to organize concepts hierarchically in an abstract form, or that they did not even have time to identify all the key concepts in the text. However, the purpose was to simulate real learning in a real-life situation in which a wealth of information had to be processed within a limited time. On this basis, it seems apparent that when participants did construct concept maps, they did not edit them extensively.

Despite some omissions most of the concepts and propositions presented in the maps, as well as their relationship to each other, were valid. This may reflect the fact that the situation lent itself to note-taking. The participants drew concept maps to summarize the content of the text while they still had the text papers. It is possible that this directed their concentration strictly to the text and did not encourage them to include their own examples or inferences in their concept maps. Another explanation for the lack of inferences may relate to the subject domain of the text and to its theoretical nature. Nevertheless, the maps were related to success in tasks requiring deep-level comprehending.

Presumably, the subject matter of the text, which concerned risk and probability theories, also limited the number of maps. The proportion of concept maps spontaneously produced by students reading a philosophical text, for instance, has been shown in earlier experiments to be considerably higher than when students are trying to learn from a text on statistics (See Lonka et al. 1994, Lahtinen et al. 1997). it may be the case that such texts do not offer as good an opportunity to exploit personal experiences as philosophical texts do. As a result, it is possible that the specific domain of statistics is not very conducive to spontaneous concept mapping in such testing circumstances as those described above.
Mental representations and science understanding

The results of this study give support to the hypothesis that the accuracy of spontaneous concept maps is related to understanding scientific text when overall success in the whole test is taken into account. It appears that participants who made average or above-average maps in terms of complexity and extent outperformed those whose maps were the least accurate. The data also suggest that drawing any kind of spontaneous concept map is superior to not using this strategy for comprehending text material. These results may be interpreted within the cognitive framework notion that the construction of mental representation is the central mediating activity between concept mapping and learning outcomes (van Dijk and Kintsch 1983, Mayer 1992, Novak 1993).

It seems likely that the activity of transforming linear text into two-dimensional graphic form requires more cognitive effort in selecting, organizing and integrating the main ideas into a compact and efficient representation of the information in the memory (Glaser 1991, Mayer 1992). Thus mapping may enhance the formation of a mental model which provides the context for understanding and making one’s own inferences about how different theories could be applied, i.e. a situation model relevant to the text.

Presumably, the extent and complexity of a map further affects the quality of situation models and consequently increases the likelihood of scientific accuracy. The findings as a whole provide support for this claim, however, viewed apart, in tasks requiring deep-level comprehension of Bayes' theorem and Propensity theory the most complete and extensive concept maps were not related to the best performance. It is possible that these theories were so demanding that applicants could not relate new information to their existing knowledge structure. In particular, Bayes' theorem has been acknowledged to be conceptually extremely difficult (see Medin and Edelson 1988). Furthermore, the lack of relevant previous knowledge might have limited participants' ability to construct situation models, regardless of the content of their concept maps. By paying adequate attention to the conceptual framework of the materials to be read, they might rather have formed a good enough textbase to give coherent answers in essay-type tasks.

Furthermore, merely including the relevant concepts of the text in a concept map proved to have only some advantage in the learning of those concepts. This result may be interpreted on the basis of Ausubel's (1968) theoretical model, according to which new concept meanings are not learned in isolation but acquired through assimilation into prepositional frameworks. Thus, in the context of statistics, comprehending initially embraces identification of central concepts, but also requires the clarification of the connections between them.

Methodological limitations

One of the methodological limitations of this study which should be mentioned concerns prior knowledge of the subject matter. For practical reasons, it was not possible to control the applicants' background knowledge of statistics. However, the main criterion for choosing the text material for the test was unfamiliarity with the domain for as many applicants as possible. Risk and probability theories were supposed not to be generally well-known to high school graduates, and in the literature, difficult even for experts (Kahneman and Tversky 1973). An attempt
was made to eliminate the potentially confounding effects of prior knowledge by writing the text with the target group in mind and using information from several sources.

It is worth pointing out that the conclusions of the study are limited to spontaneous concept maps made without revision or any improvements. Presumably, a recasting process would increase the usefulness of the worthwhile activity of concept mapping, whereas first maps in a new subject domain are usually claimed to be naive, somewhat arbitrary and asystematic (Wandersee 1990).

Obviously, the implications of the results are also limited to observable concept maps. That is, instead of drawing maps on a paper some participants may have used covert information-processing activities, for instance, linking ideas mentally or generating internal pictures (see Wade et al. 1990). It is also possible that some students have habitualized their procedures for trying to understand the text to such an extent that they do not have any need to make their knowledge structure explicit. Otherwise, these mental activities, considered to improve the quality of learning, are similar to those invoked by overt strategies such as concept mapping on a paper.

**Implications for science understanding**

Despite the methodological defects, which arose mostly from the natural learning setting, the results have implications for the use of spontaneous concept maps in science understanding and, more generally, for the need to take the quality and content of maps into account. Ultimately, students should be encouraged to construct well-organized and accurate graphic displays in a way which allows them to reconstruct the text from their own mental model, rather than reproducing what they had actually read (Kintsch 1986).

Spontaneous concept maps provide teachers with an insight into the structure and complexity of students' knowledge bases. Such utilizations thoroughly fit Pressley et al.'s (1990) suggestions that, before giving instructions about new conceptual tools for science learning, investigators should first determine how students are performing on their own. This has practical implications because it is far from easy to change students' proceduralized approach to learning (Thornton et al. 1990). On this basis, spontaneous mapping may have a stronger effect on science understanding than briefly introduced graphic tools. Equally, Wandersee (1990) argues that it may take as long as 8–10 weeks for students to become fully accustomed to concept mapping, whereas Roth and Roychoudhury (1993) recommend continued instruction to increase the quality of the structure of the maps.

What has been established here, however, is that spontaneous concept maps and their accuracy are also related to science understanding in the domain of statistics. Further, it is necessary to know under what circumstances concept mapping is truly effective. The present study reveals that spontaneous mapping is useful when deep-level understanding and application are required.

**References**


Copyright of International Journal of Science Education is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.