



The experiment consisted of rating individually each of sixty (60) concept maps. We wanted to make sure the rating of the map was based on the overall first impression of its structure and style and not on its content, so the text in the concepts and linking phrases were replaced with XXXs. Each letter in the text was replaced by an X, so the number of Xs corresponds to the length of the text (See Figure 1) In the case that the concept map had an explicit focus question as a concept, the question mark was not replaced by an X.

The experiment was carried out on a computer. Each Cmap was displayed during five (5) seconds, after which the window was cleared. During those five seconds, or after the window was cleared, the user was asked to rate the displayed Cmap as *Very Good*, *Good*, *Bad*, or *Very Bad* by clicking on the corresponding button at the bottom of the window. The subject was not allowed to go back and display a Cmap again. The subjects had to rate every single Cmap, there was no way of skipping a rating. In order to familiarize the subjects with the procedure, before the experiment began, a sample of five (5) concept maps were displayed for five seconds each. Subjects were allowed to run the sample test as many times as they wanted. The expected time to finish the experiment was five (5) minutes.

### 2.1 *The concept maps*

The concept maps that were selected were chosen to be of a large variety. Since our interest was a preliminary study to get the feeling for what variables might be interesting to study, we were not concerned with designing a precise experiment where we could obtain statistically significant results. With the results of this experiment we will design a more rigorous study on the variables that seem interesting.

A total of sixty (60) of concept maps were displayed individually to the subjects. Thirteen (13) maps had been drawn manually and forty seven (47) had been drawn using IHMC CmapTools (Cañas *et al*, 2004). Twelve (12) of the computer-drawn cmaps had resource links icons. Five (5) cmaps were displayed both in hand-drawn and computer-drawn versions. One (1) concept map was displayed twice. Two (2) concept maps were displayed in black and white, and then also displayed colored. Three (3) concept maps had the same content but they were displayed with different layout. Eighteen (18) concept maps had the focus question positioned as a concept in the upper left corner. Nine (9) cmaps had more than one color used to emphasize at least one concept.

### 2.2 *The subjects*

Because this was a preliminary study, it was not designed to produce conclusive results, but to provide a better understanding of the factors that could affect the initial (first impression) evaluation of a concept map. To make sure that there was no difference in interpretation of what a 'concept map' is, all subjects were selected because of having worked on concept mapping with Joseph Novak directly, or worked with somebody that had worked closely with Dr. Novak. The subjects were classified in two categories regarding to their experience creating concept maps. Given our personal knowledge of the subjects, we classified six (6) of them as expert concept mappers, and nineteen (19) as non-expert concept mappers. This classification was totally subjective, based on the amount of time the subject has spent 'constructing' his/her own concept maps. However, it's important to clarify that all subjects, including non-experts, have been involved with concept mapping for several years. Novak also performed the experiment.

## 3 **Results**

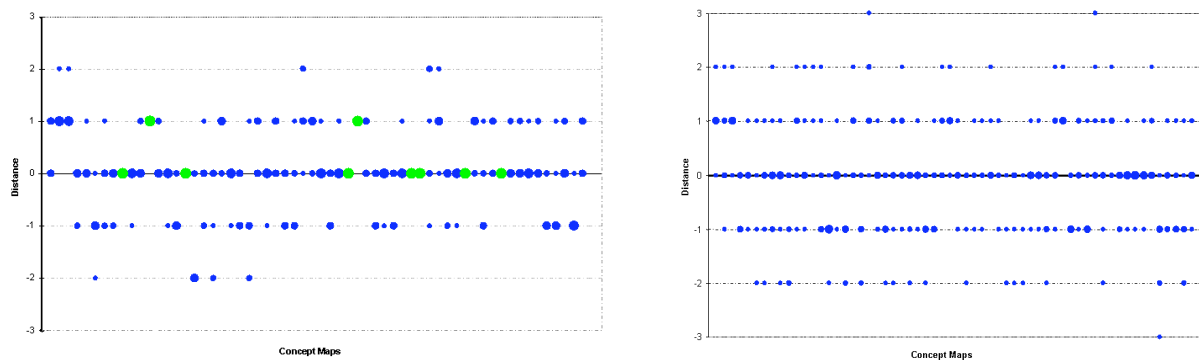
All 26 subjects completed the experiment. Although some had problems with the time it took to display the images, we feel that the results can be trusted. Each rank of a Cmap by a subject (Very Good, Good, Bad, Very Bad) received a numerical value (4, 3, 2, and 1 respectively). The overall rankings were: Very Good (7.44% of the rankings), Good (44.17%), Bad (33.78%), and Very Bad (14.62%).

### 3.1 *Experts versus non-experts*

To determine the goodness of the ranking of maps, we used Novak's ranking as the 'correct' result. We then calculated the 'distance' of each ranking from Novak's ranking as the numerical difference between the two. A distance of zero meant that the subject agreed with Novak in the ranking of that particular Cmap. Figure 2 shows a

plot of the distances for the expert subjects on the left, and for the non-experts on the right. In each graph, the x-axis is the number of the Cmap (total of 60) while the y-axis shows the rankings by the subjects. The size of the dots is proportional to the number of rankings that have that particular distance. The larger dots show the cases where all subjects agreed (9 Cmaps for the experts, none for the non experts). The graphs clearly show that experts tend to agree with Novak more often than non-experts.

As a group, all the expert subjects agreed with Novak in the ranking of seven (7) Cmaps; for two other Cmaps they agreed as a group, but the Cmap was ranked lower than Novak's evaluation. The non-experts never agreed as a group on a ranking. None of the expert subjects had an absolute distance greater than two with Novak for any ranking. In 4.44% of the rankings, the absolute distance between Novak and the experts was two, 40.83% was one, and 54.72% was zero, meaning that in 54.72% of the rankings an expert subject agreed with Novak. This means that in 95.55% of the rankings experts were at most a distance of one from Novak's. For five rankings (0.44%) the absolute distance between Novak and a non-expert subject was three, the largest possible value. The non-expert subjects agreed in 43.77% of the rankings with Novak, had a distance of one in 46.05% of the rankings, and a distance of two in 9.74% of the rankings. In 23.68% of the rankings the non-expert users ranked the cmap lower than Novak and 32.54% they gave a higher ranking than the one conferred by Novak. The expert subjects, ranked 25.28% of the rankings lower than Novak and in 20.00% of the rankings it was higher.



**Figure 2.** Distance in the ranking of the Cmaps between Novak and the expert subjects (left graph) and non-expert subjects (right graph).

Experts were on the average faster than non-experts in evaluating the concept maps (nine minutes nine seconds versus eleven minutes two seconds). The average response time to finish the experiment was 10 minutes and 42 seconds.

### 3.2 Concept map features

*Number of concepts:* Concept maps with less than six concepts received an average rank of 2.10, and concept maps with more than twenty concepts had 2.74 on average. The number of concepts per cmap varied between 4 and 61.

*Hand drawn versus computer drawn:* Five (5) cmaps were displayed in both hand-drawn and computer-drawn versions, as shown in Figure 1. The computer-drawn versions were created using IHMC CmapTools. For two pairs of Cmaps, 65% of the subjects gave the same ranking to the hand-drawn and computer-drawn versions, and for the other three pairs of cmaps the same value was given in 92%, 38%, and 31% of the rankings. In all the five pairs, the computer-drawn versions received an overall better evaluation than the hand-drawn cmaps. The computer-drawn versions had 2.05 on average and the hand-drawn versions 1.81.

*Repeated map:* One concept map was displayed twice, as the first map and in the 48th place. The second time it was displayed, 23.0% of the subjects ranked it better, 57.69% gave the same ranking, and 19.23% provided a lower ranking.

*Coloring:* Two pairs of cmaps were displayed with and without colors. In both cases the colored cmap was displayed before the black and white cmap. For the first pair, the cmap with colors was ranked better by 3.85%, the same by

80.77%, and lower by 15.38% of the rankings. For the second pair, 19.23% ranked better the concept map with colors, 57.69% the same, and 23.08% lower.

*Layout:* Three concept maps with the same content were displayed with different layouts; in two cases the main transformation was to position the root concept in a different location and in the other case the concepts were distributed horizontally. In the first pair, the version that was displayed first had the root concept in the center of the cmap, while the version displayed later had the root concept at the top showing a hierarchical layout. The second cmap displayed was ranked the same in 53.85% of the rankings, 30.77% better, and 15.38% worst. For the second pair, the first cmap displayed had the root concept on the top of the cmap, and the second had the root concept in the center of the cmap. It was ranked similar by 34.62%, better 11.54%, and 53.85% worst. In the last pair of cmaps the concept map with a less dense layout was ranked better by 30.77%, same by 57.69%, and worst by 11.54% of the ranking than an equivalent map with a denser layout.

#### **4 Discussion**

It's important to recall that none of the results of this experiment have statistical validity; all results are based on a small number of items and are to be used in the design of a more robust experiment.

The results seem to indicate that expert concept mappers tend to agree more when assessing a concept map that is presented for a short period of time and without being able to examine its contents. Figure 2 supports this, showing that there is a lot less variance in the assessment by the experts with respect to Novak's assessment than in the assessment by non-experts. This indication is in accordance to the expertise literature that reports that experts perceive meaningful patterns that allow them to make decisions (Cooke, 1992). The results also seem to indicate also that non-experts tend to assign higher rankings to concept maps than Novak or the experts group. The results suggest that experts are faster than non-experts in performing the assessment tasks at the same time that they are commit less errors (in this case, are closer to Novak's rankings), which is accordance to what Cooke (1992) reports.

The results on the variations on the concept map features show that cmaps with a small number of concepts (less than six concepts), received the lowest ranking (on average) and those with more than twenty concepts were ranked better. It seems that the number of concepts influences the first impression on the quality of the map. The results suggest that hierarchical concept maps with the root concept on the top are ranked as better maps than exactly the same map with the concept located in the center. The experiment seems to suggest that that concept maps that are computer-drawn are perceived to be better than the same map hand-drawn. The experiment did not provide evidence that the presence of colors in a concept map affects the subjects' first impression.

#### **5 Conclusions and Future Work**

This preliminary study seems to indicate that there is merit in continuing this line of research on how concept maps are perceived by expert and non-experts, and whether topological and style features in the concept map affect their initial assessment. As concept maps themselves are commonly used for assessment, a more conclusive study may provide results of importance to the concept mapping community.

#### **References**

- Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M., & Carvajal, R. (2004). CmapTools: A Knowledge Modeling and Sharing Environment. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the 1st International Conference on Concept Mapping*. Pamplona, Spain: Universidad Pública de Navarra.
- Cooke, N.J. (1992). Modeling Human Expertise in Expert Systems. In Hoffman, R. (Ed.). *The Psychology of Expertise, Cognitive Research and Empirical AI*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Novak, J. D., & Gowin, D. B. (1984). *Learning How to Learn*. New York: Cambridge University Press.
- Ruiz-Primo, M.A. (2004). Examining Concept Maps as an Assessment Tool. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the 1st International Conference on Concept Mapping*. Pamplona, Spain: Universidad Pública de Navarra.