QUALITY OF MAPS - MEASURING COMMUNICATION

Senior scientist, Dr. Lars Brodersen
National Survey and Cadastre - Denmark (KMS)
Rentemestervej 8
2400 Copenhagen NV
Denmark
fax: +45 35875052
e-mail: lrb@kms.dk

Senior scientist Hans K. Andersen
Senior scientist Steen Weber
Risø National Laboratory
4000 Roskilde
Denmark
fax +45 46775199
e-mail: hans.andersen@risoe.dk, steen.weber@risoe.dk

Abstract

How can we assess quality of the cartography of a map with regard to communicating map information to the user? How can we assess the quality of achievements and product development if we can't measure the quality? Map producers may need to carry out quality measurements of map prototypes before release to users as it is common practice in other industries, e.g. a car manufacturer carries out 'customer clinics'; new drugs are tested in user trials in order to test their effectiveness.

A series of tests were carried out with the purpose of developing and assessing a methodology for measuring the quality of communication on maps. The aim was to develop a standard method, which can be used to test maps, map prototypes, and other forms of geographical information, in large-scale tests with representative user groups.

The method consists of four parts: 1) tasks to be solved with the aid of a map, 2) eye tracking recording, 3) think aloud method, 4) interview after completing the tasks.

The study was based on ten subjects/users with no specialist map knowledge. These subjects were asked to answer 22 questions with the help of two real topographic maps that differed in design but covered the same area and had the same content. A video camera was used to observe the subjects and record their answers and verbalised thought processes. Eye movements were recorded using a helmet-mounted eye-tracking device.

The result was a measure of the time taken to solve each task, number of words used and whether the answers were correct or not. The eye movements were used to assess the type of fixation, length and number of fixation and saccades. Marks were also given for time taken, strategy and overall performance. Correlation between quantitative factors was sought. The different measurements used to assess the subject also acted as a 'control' to highlight conflicts between what was said by the subject and what the subject did. Uniform/similar behaviour or statements by the subjects were also noted.

Part of the study was for example to investigate the number of eye fixations between map and legend to see if one map design requires more consultation in the legend than the other design.

The results were in general remarkable. Overall answer correctness (based on 220 answers) was 60% with standard deviation 4%. Time used per question averaged 78 sec with a standard deviation 11 sec. In the markings for performance (66 marks per subject given by the researcher), the standard deviation was also 4%. Other results show the same clear, uniform tendency. The overall conclusion was that the method is useful because the performance of the subjects was uniform.

What is map reading?

When designing a research about map reading it is essential to define what is understood by 'map reading'. We have chosen an *operational definition* rather than a theoretical. As a consequence of this, the research is carried out on the basis of real map examples and questions asked to those maps. 'Map reading' is defined as the necessary in order to answer the questions, to solve the tasks.

What interest us most are the subject's skills in realistic, daily-like map reading of real maps. Whether these skills belong to one or another theoretical group of behaviour is another thing - interesting for another research.

One could say this project uses a ***functional, communication oriented*** definition of map reading. The operational definition of map reading has the impact that the choice of maps and questions is crucial. In principle, the results of this project will only be valid for these maps in conjunction with these questions. This does not mean that the results are of limited interest. The operational definition does not exclude considerations of what *skills* are needed to solve the individual tasks. Some tasks may imply the subject to be able to recognize the single signatures only without being able to interpret them, e.g. on a street atlas. Other tasks may need more than that, e.g. relation to other signatures, integration of all signatures in an area, conjunction to general knowledge outside the map. When we know what skills are needed to solve specific tasks, it should be possible to generalize and extrapolate the results to other maps and tasks demanding same skills.

Another important point is to define what is understood by 'good map reading'. The ultimate aim for the cartographer is that his map affords the user the possibility of quickly and certain getting answers to relevant questions. This definition does only apply to 'structured information search' with a given aim. It does not apply to a so-called 'encyclopædia search' where it is not known from the beginning what the aim is. The key-words from the definition are 'quickly', 'certain' and 'correct answers'. The appropriate measures in an attempt to quantify these three terms could be: 1) time used to solve the task, 2) visual behaviour identified through observation of the eyes, or observation of the subject, and 3) number or percentage of correct answers.

Eye-movement recording

In the early 1970s, cartographers became very much interested in examining the map-reading process, applying eye-movement tracking methodologies (see e.g. Muerhcke, 1972). They hypothesised that such studies could produce new knowledge to support the process of symbol design, and give cartographers more control over how people read their maps, and to create more usable and efficient maps. Still, we have not yet seen the fulfilment of the hypothesis. One reason is that these studies mainly focused on 'where' people look in the maps not 'how' they look (Castner and Eastman, 1984; 1985). Several studies within other areas of research (on picture perception) have focused exactly on that problem.

Recently eye movements tracking has come to play an important role in computer interface evaluation (Airbus airplanes, Danish Railways etc.) (see e.g. Andersen, 2000 and Weber, 2000). We do not se much methodological differences in evaluating maps compared to evaluating conventional computer user interfaces. Both are ‘maps’ that contain ordered spatial symbolic information. Goldberg and Kotval (1998) has proposed a series of measures for evaluating user interfaces based on eye-movement tracking. It’s a list of content independent quantitative eye-movement measures divided into temporal and spatial dimensions. These measures have to be analysed in relation to the tasks given a subject during a usability test.

Fixation duration (unit: milliseconds). Longer fixation indicates the person is spending more cognitive processing time to interpret symbolic information.

Number of fixations (unit: number). Related to a low or high degree of cognitive efforts in searching meaningful objects.

Scanpath duration (unit: milliseconds) and **length** (unit: millimetres). Time spend in area of interest (AOI). Can be used to compare norm (optimal) visual behaviour with observed behaviour, i.e. does the person follow a sensible and rational strategy?

Number of saccades. A large number indicates low degree of search efficiency and poor interface layout. In this context 'saccades' means jumps between dwells on AOI.

Saccadic amplitude (unit: millimetres). Used together with other measures the amplitude level can be an indicator of search complexity.

Post target fixations or **backtracking:** If the person goes back to an already visited target it indicates the person’s doubtfulness regarding the target’s meaningfulness.

Blinking frequency (unit: number per second). High frequency indicates uncertainty. Low frequency indicates resolution.

The crucial point is the correlation between the above measures and the results from 'instructor's rating' (performance data). If a 'good result' cf. above measures is confirmed by a 'good result' cf. 'instructor's rating', it is very likely our method is working according to hypothesis (however, more work still has to be done to refine the method). If correlation is low, more work has to be done to make a new starting point for the method.

Methods and experimental set-up

We have used a ‘SMI iView’ eye-tracker (http://www.smi.de/iv/index.html) combined with a ‘Polhemus head-tracker’ (http://www.polhemus.com). We used the ‘SMI HeVisual software’ for monitoring validity of the six dimensional data during recording; X,Y, and Z for the position of the head; azimuth, elevation, and role for the heads rotation, as well as the time. Output from the software is: X, Y on the map, time, distance between subject and map, as well as the three rotation parametres of the head. One data file was created per question per subject.

We asked the subjects to ‘think-aloud’ during the test. The speak was recorded as audio data on a scene-video; Sony DV-Cam. One videotape per subject all questions. Moreover the scene-video of subject and map during the test, was used to analyse non-verbal behaviour like pointing, nodding etc.

All videos were manually analysed. Every word, said by the subject, were written down. At the same time it was recorded how much time the subject had spend answering each question, through time code recording on the video. From this, it was extracted what answers the subject had given. Finally, the adequacy of applied strategy to solve problems was assessed though wording, physical behaviour, pointing, retries, search-strategy etc.

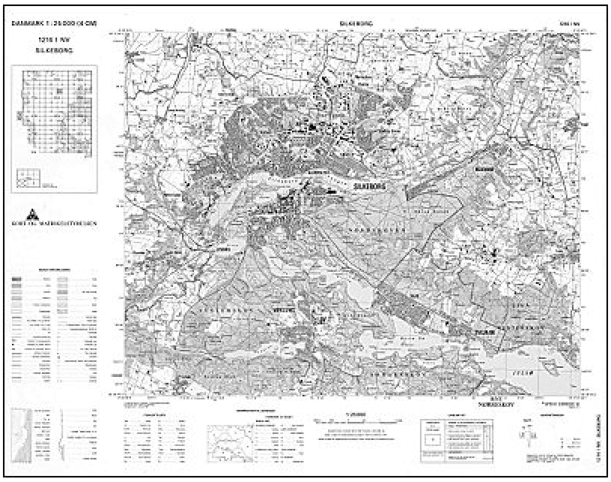
Procedure, subjects and material

Ten subjects participated in the experiment, five females and five males. The age of the subjects ranged from twenty-three to sixty-one years. They all had a high level of education. They were four senior scientists, three marketing staff members and three computer technicians (i.e. used to complex, abstract thinking). Two authentic topographic maps from National Survey and Cadastre Denmark were used. We created two map-reading tasks containing eleven questions each. Each test lasted approximately one hour. Three researchers took part in the test.

The following procedure was used. Firstly, the subjects were given an introduction to goals and content of the study. Secondly, the experimental set-up was explained in terms of equipment. Thirdly, instruction was given how to handle the tasks during the test. It was emphasised we were testing usability of the maps and applicability of the method, not the subject’s skills.

After this introduction the subjects were given a small questionnaire (regarding their demographic data). Next step was calibration of the head- and eye-tracking device, plus calibration of the scene camera add on. Each question was written on a paper and handed over to the subject one by one. The subject was asked to read the question aloud (to check if they actually understood the text) and then start answering. It was made expressively clear to the subject that they could ask anything they wanted (to simulate a daily like map reading situation as much as possible). When the subject had finished first task, the eye-data were saved and calibration verified. Then the subject went on with second task.

The maps



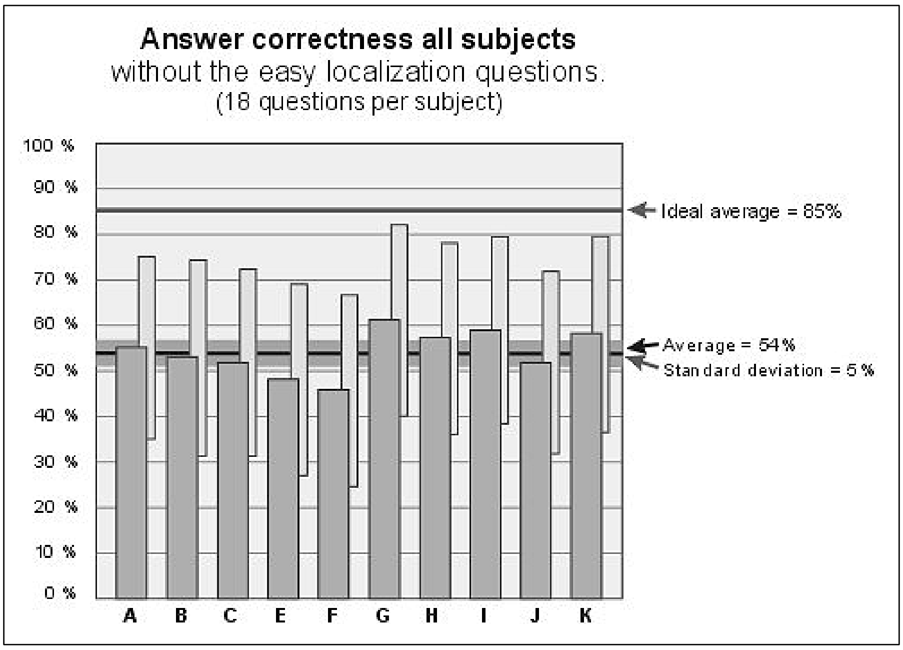
For this particular research project, we chose one topographic map available in two designs (only one is shown here). Purpose of the project was to assess the method and *not* to assess maps as such. When the method one day has been fine-tuned and is being used in product development activities then, of course, it is crucial to work thoroughly with map design and the choice of maps. The intention behind the difference in the map design between the two versions was to increase map reading efficiency (faster, safer and more directly access). Therefore, we found it obvious to use this map in two versions as test material when assessing a method to measure map-reading quality.

Questions

A map is not a task in itself. However, when a map is connected with expectations of a certain profit, one can talk about a map reading task. Our aim was to assess the *practical* reading ability with standard maps, not to assess reading ability based on a theoretical standard. It is pretty well defined what different types of maps can be used for; topographic maps to find information related to things on the earth's surface etc. Therefore, the chosen questions should reflect standard use of the chosen maps. In addition to this, the questions in total should cover essential, standard reading skills:

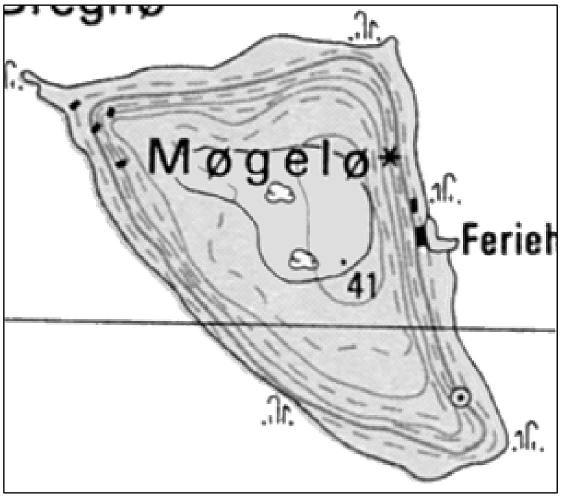
- Localization of information (e.g. '*Point out the legend'*).
- Decoding, i.e. identification and understanding of single symbols (signatures) (e.g. '*Point out a stone fence'*).
- Literally understanding, or combined decoding, i.e. understanding of the openly formulated (wording) (e.g. '*What is the height difference to walk from the lakeside at Sejs to the top of Sindbjerg?'*).
- Interpretative understanding, i.e. understanding not directly formulated context (e.g. '*Is there, has there been or has there never been a railway to Skægkær?'*).

Results:



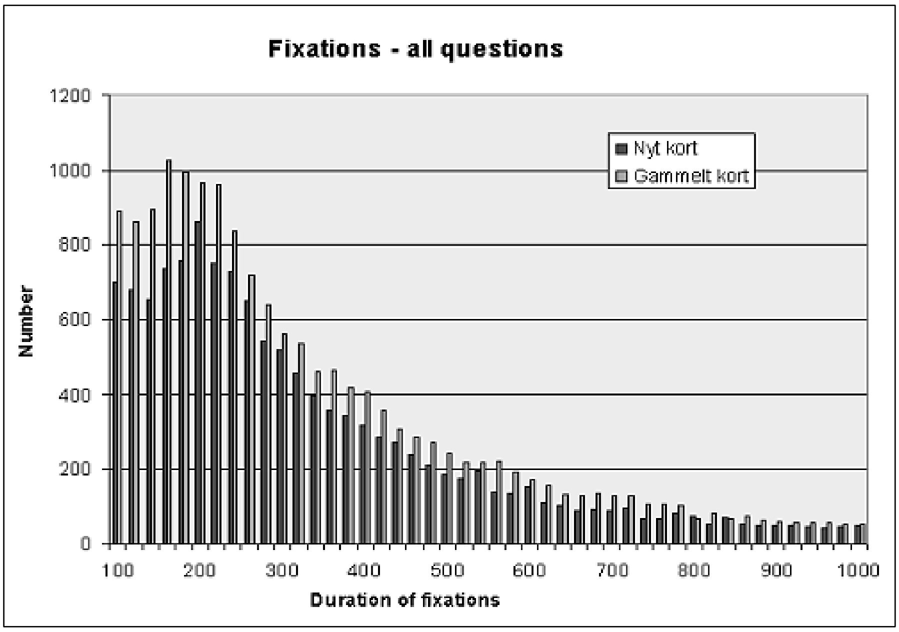
If we exclude the easiest localisation questions the test showed that the subjects were able to answer correct on half of the questions. This is way below the expectations of National Survey Denmark to their customers. This might, by the way be the most important result why we must carry out research in this field. It is also important to notice the uniform results; the subjects did not behave and answer identically. Far from. But the overall result is the same for all of them.

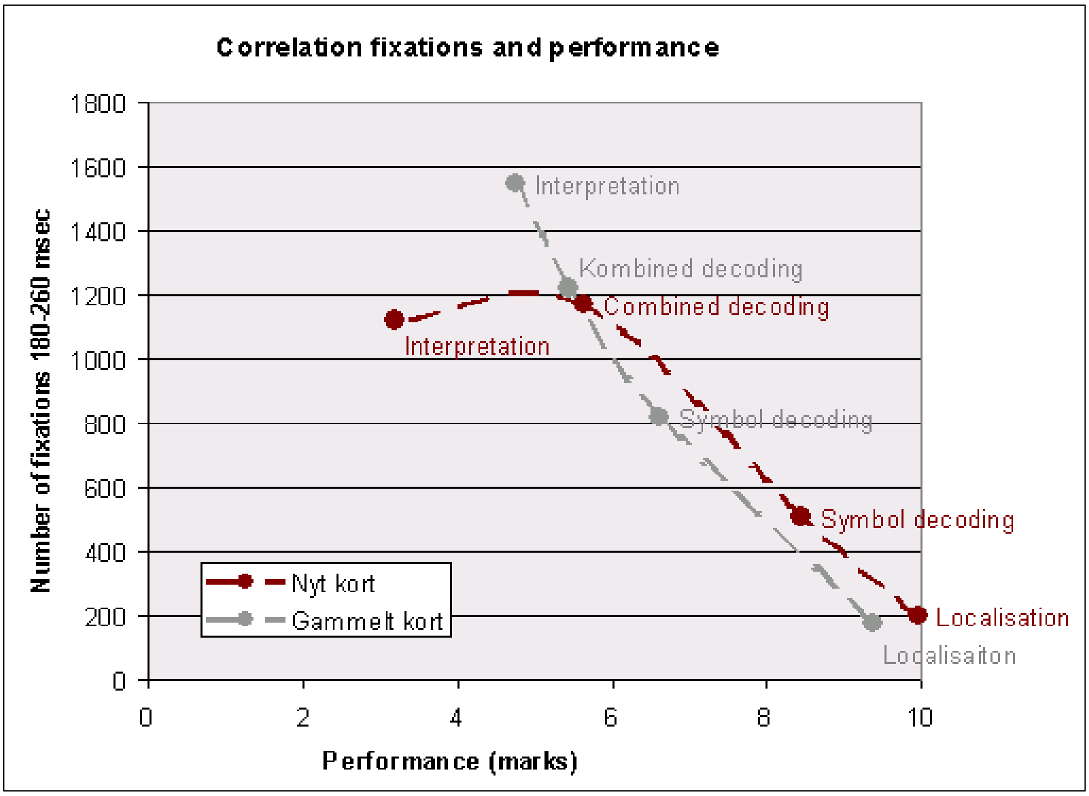
One questions was:
<u>16. On Møgelø there are five different point symbols. What do they stand for (each of them)?</u> *(answer: Hardwood, conifer, memorial, spot height, building).*
Results: Three out of ten subjects identified all five point symbols, three subjects identified four symbols, one subject identified three symbols, two subjects identified two symbols and one subject identified no point symbols.
- It was observed that all subjects identified one symbol per visit in the legend. Obviously there is only room for one symbol in the short-term memory.



This figure shows the frequency of fixations as number of fixations versus duration of fixation. There is an overweight of short fixations in the old map compared to the new map. On the old map all subjects perform a higher number of fixations in all duration categories.

Solving the tasks pertaining to the old map simply requires more fixations of every category. The plot is skewed with a peak of fixations falling into the range of 100-300 msec. This indicates that fixation duration between 200 and 250 msec reflects a high level of heavy cognitive work. We also see that the overweight of short fixations, which are characteristic for looking behaviour with respect to reading and picture viewing, also seem to characterise map examination.





The figure above shows the correlation between fixation and performance. We still have to do calculations on this item. However, the figure indicates a clear tendency between the eye's fixation behaviour and the performance. If this was not the case, the points on the diagram would have been placed 'randomly'.

Conclusions

The number of fixations in a map and the apparent amount of information in that area of the image seem to be useful for evaluation of maps. That is, a large number of fixations indicate high complexity. However, a certain caution must be exercised when dealing with this conclusion. A conclusion like this requires full insight into the relation between a task and its needs for a certain visual behaviour, which we don't have at the moment. In order to get closer to this insight, we need to examine our task related data (performance data) in detail and relate these more systematic to the visual data.

The duration of fixations seems to indicate the level of map complexity as well. That is, longer fixations indicate more cognitive processing, which should be equal to a high level of map complexity. In this study we did not find much difference between the two maps with respect to duration of fixations. One could consider that the two maps to some degree share complexity, but of a different kind.

In addition, the results seem to indicate that the map reading tasks given to the subjects were too complex or difficult. We have to look into the complexity of tasks again and consider using simpler tasks to see whether this will have an effect or not. It looks very much like we overrated the level of map reading abilities, which people in general have. It is remarkable that the subjects were able to answer correct on approximately half of the questions. Expectations among the staff at National Survey and Cadastre was, until this study, that this kind of users (high education and average map user experience) could answer 85% correct, to these ‘normal’ questions.

In general the instructor ratings show very small standard deviation. This could indicate that the subjects’ behaviour, created by the experimental set-up, is normalized. If all subjects behave approximately equal the results should be the same whether we examine many subjects or just a few. The equality is not caused by monotonous rating by the instructor, which can be see from the standard deviation on the ratings.

Independent from the eye movements and instructor ratings, the sheer observation of the subjects and the examination of their answers reveal useful information regarding map communication.

References

Andersen, H., H., K., and Hauland, G.: Measuring Team Situation Awareness of Reactor Operators During Normal Operation: A Technical Pilot Study, In *Proceeding of the HPSAA Conference*, pp.268-273, 2000

Brodersen, L.: Kort som kommunikation – teori og metode i kartografien. Forlaget Kortgruppen a/s, 1999. ISBN 87-984113-1-4.

Castner, W., Henry and J., R., Eastman: Eye-Movement Parameters and Perceived Complexity - I in *The American Cartographer*, Vol. 11, No. 2, 1984, pp. 107-117.

Castner, W., Henry and J., Ronald, Eastman: Eye-Movement Parameters and Perceived Complexity - II in *The American Cartographer*, Vol. 12, No. 1, 1985, pp. 29-40.

Goldberg, J.H. and Kotval, X., P.: Eye Movement-Based Evaluation of the Computer Interface, in Kumar, S.K. (Ed.), *Advances in Occupational Ergonomics and Safety*, Amsterdam: IOS Press, 1998, pp. 529-532.

Hauland, G.: Building a Methodology for Studying Cognition in Process Control: a Semantic Analysis of Visual and Verbal behaviour, *Main thesis in psychology for the degree of Candidatus Rerum Politicarum*, The Norwegian University of Science and Technology, 1996.

Muehrcke, P., C.: *Thematic cartography*, Association of American Geographers, Resource Paper, No. 19, 1972.

Weber, Steen, T., Bove, C. Le Draoullec, F., Reuzeau, H., B., Andersen: Pilot's response and eye fixations during take off. *Risø National Laboratory*, 2000.