## Original Article

# Measuring effectiveness of graph visualizations: A cognitive load perspective

Weidong Huang[a],*
Peter Eades[b] and
Seok-Hee Hong[b]

[a]National and International Surveys Program, Australian Council for Educational Research, Camberwell, VIC 3124, Australia.
[b]School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia.

*Corresponding author.

**Abstract**    Graph visualizations are typically evaluated by comparing their differences in effectiveness, measured by task performance such as response time and accuracy. Such performance-based measures have proved to be useful in their own right. There are some situations, however, where the performance measures alone may not be sensitive enough to detect differences. This limitation can be seen from the fact that the graph viewer may achieve the same level of performance by devoting different amounts of cognitive effort. In addition, it is not often that individual performance measures are consistently in favor of a particular visualization. This makes design and evaluation difficult in choosing one visualization over another. In an attempt to overcome the above-mentioned limitations, we measure the effectiveness of graph visualizations from a cognitive load perspective. Human memory as an information processing system and recent results from cognitive load research are reviewed first. The construct of cognitive load in the context of graph visualization is proposed and discussed. A model of user task performance, mental effort and cognitive load is proposed thereafter to further reveal the interacting relations between these three concepts. A cognitive load measure called mental effort is introduced and this measure is further combined with traditional performance measures into a single multi-dimensional measure called visualization efficiency. The proposed model and measurements are tested in a user study for validity. Implications of the cognitive load considerations in graph visualization are discussed.
*Information Visualization* (2009) **8**, 139–152. doi:10.1057/ivs.2009.10

## Introduction

It is widely accepted that visualization assists people in understanding data and communicating information about the data with each other, by providing 'cognitive support'[1] through a number of mechanisms. For example, representing data visually makes it possible for some tasks to be done by using a few simple perceptual operations, which otherwise can be a laborious cognitive process; a visualization serves as external memory to reduce demand on human memory.[2]

However, this does not mean that visual representations are always more effective and less demanding than data in their original format. Pictures themselves cannot automatically improve performance of human understanding.[3] People process information in working memory[4]; working memory has a limited capacity. An inappropriate visualization may impose high cognitive load on the human cognitive system, thus overwhelming the viewer and undoing its benefits of cognitive support. For example, in performing a complex task with line graphs of the same data, Lohse[3] found that subtle changes in line graphs (for example,

graph legend) may cause cognitive overburden degrading task performance significantly.

In graph visualization, algorithms are generally developed to draw graphs according to some pre-defined aesthetic criteria, based on the assumption that the resulting pictures are easy to read, easy to understand and easy to remember. However, success of visualization largely depends on whether the viewer can perceive and process the embedded information effectively.[1,5] Unfortunately, in designing graph visualizations, people heavily rely on general conventions or on their own insights.[6] Although calls and attempts of looking into theories from general psychology have been made from leading researchers,[7–9] the role of human factors such as the limitation of working memory on visualization understanding has not yet been fully explored.[1] Herman *et al*[10] summarize this situation nicely as follows:

> ... papers in information visualization rarely apply cognitive science and human factors. This is for no lack of trying; very few of the findings in cognitive science have practical applications at this time, and very few usability studies have been done. Cognitive aspects are undoubtedly a subject for future research. For this reason, an objective evaluation of the merits of a given approach is difficult.[10]

In previous evaluations of graph visualization, performance measures including response time and accuracy are commonly used to measure effectiveness (for example, Purchase[6] and Ghoniem *et al*[11]). While it is claimed that a visualization is more effective if it allows the viewer to take less time to complete the task and make fewer errors, it is not often that a visualization can result in a high accuracy and a short response time together.[12,13] Instead, it is likely that the viewer obtains a relatively higher accuracy, but at the same time takes more time, largely because of the widely acknowledged speed-accuracy tradeoff. It is this inconsistency between individual measures that makes it difficult to judge the overall quality of the visualization, thus offering little practical guidance on whether one visual representation should be preferred over another.

Further, when working with, for example, two visualizations of the same data (suppose that one of the two is bad and the other is good), it is feasible that the same viewer expends more mental effort to compensate for the increased cognitive load induced by the bad visualization, thereby maintaining the same level of performance as with the good one. In this case, performance-based measures alone, therefore, are not sensitive enough to detect the difference in the amount of cognitive effort devoted with the two visualizations.[14]

To summarize, in evaluation of graph visualizations, we need a new measure to discern the difference in the amount of cognitive effort devoted. To make the judgment of the overall visualization quality easier, it is also necessary to unify all individual measures into a single one in a meaningful way.

As an initial attempt to meet the above needs, we explore the application of cognitive load considerations to graph visualization. In particular, the effectiveness of graph visualizations is measured from a cognitive load perspective, based on recent results of research on Cognitive Load Theory (CLT).[15] Our main aims are

1. To introduce two measurements for evaluation of graph visualizations: mental effort and visualization efficiency.
2. To propose a cognitive model to reveal the interactive relationships among task performance, mental effort devoted during performance and memory demand induced by the task.
3. To conduct a user study to validate the proposed model and cognitive measurements.

## Background

### Information processing model of memory

Human memory can be considered as an information processing system. It has three basic components: *sensory memory*, *short-term memory* and *long-term memory*.[16] As can be seen from Figure 1, the incoming information from outside must be processed through sensory memory. Attended information is then passed to short-term memory and processed in working memory.[4] To remember a piece of information, efficient cognitive strategies are required for transferring the information from sensory memory all the way to long-term memory where it can be stored. The information stored in long-term memory can also be retrieved back to working memory for processing when it is needed.

The bottleneck for efficient information processing is that working memory has its limitations in *duration* and *capacity*. Working memory has a finite capacity, known as 'the magical number seven plus or minus two'.[17] Moreover, information stored in working memory decays over time.[18] Apart from information storage, working memory is also responsible for processing information elements during the performance of cognitive tasks. This means that there is considerable competition for working memory. When elements to be simultaneously processed are highly interactive, the number of elements and the time period of being held in working memory can be further dropped. A direct implication of these limitations is that if a cognitive activity requires (1) many elements to be held in working memory, (2) many elements to be processed simultaneously (which is called high 'element interactivity'[15]), memory overload and information loss can be expected, resulting in poor performance.

### Cognitive load theory

Cognitive Load Theory emphasizes the limitations of working memory, and considers cognitive load
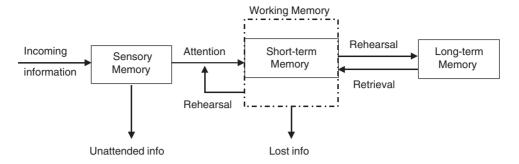
**Figure 1:** Information processing model of memory.

associated with the learning of complex cognitive tasks as the major factor determining the success of instructional methods.[19] The theory suggests that for better learning results, instructional materials should be presented in alignment with cognitive processes. Cognitive load induced by instructional formats is not relevant to learning and should be reduced, so that more memory resources can be available for knowledge acquisition. Cognitive load can be controlled by manipulating instructional formats (for an overview, see Van Merrienboer and Sweller[15]).

*Measuring cognitive load*
*Cognitive load* refers to the amount of cognitive resources needed to perform a given task; therefore cognitive load can also be called 'memory demand'.[20] *Mental effort* refers to the amount of cognitive capacity that is actually allocated to accommodate the demand imposed by the task; thus mental effort reflects the actual cognitive load.

Techniques used to measure mental effort can be classified into three categories (for a review, see Paas *et al*[14] and De Waard[21]): subjective (or self-report) measures such as rating scales, performance-based measures such as secondary tasks and physiological measures such as pupillary dilation. Each of the measuring techniques has its own advantages and disadvantages. In particular, rating scale techniques assume that 'people are able to introspect on their cognitive processes and to report the amount of mental effort expended'.[14] This assumption may appear questionable, however it has been demonstrated that people are quite capable of indicating their perceived mental burden by giving a numerical value.[22] Paas *et al*[14] reviewed previous studies that used different types of measuring methods, and further concluded that rating scale techniques 'are easy to use; do not interfere with primary task performance; are inexpensive; can detect small variations in workload (that is, sensitivity); are reliable; and provide decent convergent, construct, and discriminate validity'.[14] And indeed, the subjective rating scale measure has remained the most used technique among researchers.[23]

There are two main types of subjective rating scale techniques available in the literature. The first type uses multi-dimensional rating scales, in which overall mental effort is obtained by combining the assessments of a number of factors used to describe individual aspects of mental effort. Examples include the NASA Task Load Index (TLX) and the Subjective Workload Assessment Technique (SWAT). The second type measures mental effort based on a univariate scale, in which a single number is used to represent the overall mental effort.

In the context of CLT, Paas[24] developed a univariate 9-point Likert scale. The subject is required to report mental effort expended by mapping the perceived amount of mental effort to a numerical value. The numerical values range from '1' to '9', corresponding to verbal labels from 'very, very low mental effort' to 'very, very high mental effort'. There are also modified versions that have been used in cognitive load research since the Paas scale, such as the 7-point scale of Marcus *et al*.[25]

Specific research has been done to compare the two types of rating scale techniques,[23] and to compare the scale of Paas and its variants.[14,26] The results suggest that for overall mental effort, the original Paas scale is reliable, non-intrusive and sensitive to small changes in memory demand.[26]

*Measuring instructional efficiency*
Given the complex relation between cognitive load and performance that is often mediated by human factors such as motivation and limited working memory, measured cognitive load is meaningful only when it is interpreted in association with its corresponding performance level.[14] To compare relative efficiency of instructional conditions (E), Paas and Van Merrienboer[27] developed a computational method. This method combines mental effort expended (for example, testing effort) and performance attained (for example, test scores) during the test phase by standardizing raw performance and mental effort data to corresponding *z*-scores and using the equation:

$$E = \frac{z_{\text{Performance}} - z_{\text{Testing\_effort}}}{\sqrt{2}} \qquad (1)$$

In Equation (1), mental effort is related to performance in the context of instructional efficiency. High instructional efficiency is achieved when high task performance

is attained in association with low mental effort, and *vice versa*.

To be able to measure instructional efficiency more precisely, Tuovinen and Pass[26] extended the above two-dimensional method by adding a third dimension measure, mental effort during the learning (training) phase (for example, learning effort) into the equation:

$$E = \frac{z_{\text{Performance}} - z_{\text{Learning\_effort}} - z_{\text{Testing\_effort}}}{\sqrt{3}} \quad (2)$$

For more details about the equations, see Tuovinen and Paas[26] and Paas and Van Merrienboer.[27]

## Visualization Efficiency

Based on the review above, mental effort can be used to measure the amount of devoted cognitive effort. To combine mental effort and performance measures in a meaningful way, these measures are categorized into two groups: cognitive gain and cognitive cost. Then *visualization efficiency* can be defined as the extent of the cognitive gain relative to the cognitive cost. In this context, an equation for visualization efficiency can be derived using a similar approach to that for instructional efficiency, which is outlined as follows:

Given data sets for three dependant variables: response accuracy (RA), response time (RT) and mental effort (ME), since they are measured in different types of units, we need to standardize them into $z$ scores first to make them comparable.

To be more specific, suppose that there are $n$ experimental conditions and each condition yields $m$ data entries for each of the dependant variables, that is, $n \times m$ entries for RA, $n \times m$ entries for RT and $n \times m$ entries for ME. To standardize each entry for RA, let $\mu$ be the grand mean and $\sigma$ be the standard deviation of the $n \times m$ entries. The $z$ score of a raw entry $X_{ij}$ (where $i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$) can be computed by using the equation:

$$z_{\text{RA}} = \frac{X_{ij} - \mu}{\sigma} \quad (3)$$

The same procedure can be used to compute $z_{\text{RT}}$ for RT and $z_{\text{ME}}$ for ME. For our purpose, the $z$ score can be understood as a common yard stick for the data entries. A $z$ score indicates how far (measured in standard deviations) and in what direction $X_{ij}$ deviates from the grand mean.

Then we obtain visualization efficiency (E) using the equation:

$$E = \frac{z_{\text{RA}} - z_{\text{ME}} - z_{\text{RT}}}{\sqrt{3}} \quad (4)$$

In this equation, mental effort is related to performance measures in the context of visualization efficiency, which

is defined as the difference between cognitive cost (ME plus RT) and cognitive gain (RA). High efficiency is achieved when high accuracy is attained in association with low mental effort and a short response time, whereas low efficiency occurs when low accuracy is associated with high mental effort and a long response time. When E = 0, the cognitive cost and performance accuracy are balanced.

Note that $\sqrt{3}$ in Equation (4) makes it possible to graphically illustrate E in a three-dimensional Cartesian space. If three axes are drawn with the vertical axis representing RA and the other two representing ME and RT, respectively, on a horizontal plane, and the $z$-scores of the three measures are plotted as a point in the space, then E represents the perpendicular distance from this point to the plane whose equation is

$$z_{\text{RA}} - z_{\text{ME}} - z_{\text{RT}} = 0 \quad (5)$$

## Model of Task Performance, Mental Effort and Cognitive Load

### The construct of cognitive load

Paas and Van Merrienboer[28] consider cognitive load as a multi-dimensional construct that represents the load imposed on human memory during performance of a cognitive task. The construct has a causal dimension representing factors that may affect cognitive load, and an assessment dimension representing factors that may be affected by cognitive load.

Based on this framework, we propose a conceptual construct of cognitive load in the context of graph visualization, which is illustrated in Figure 2.
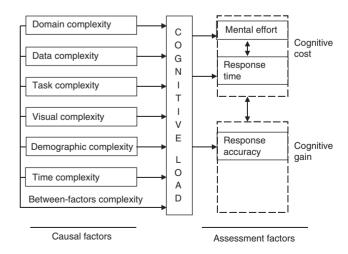


**Figure 2:** The construct of cognitive load for visualization understanding.

*Casual factors*

1. *Domain complexity*: Different application domains have different data formats and contents, define different tasks, and require different visualizations. This factor defines specific visualization constraints and requirements, and provides a basis for overall complexity. For example, visualizing biological data should be different from visualizing social network data, because of the difference of information that biologists and sociologists look for and the distinction of the data in nature.
2. *Data complexity*: This includes the number of objects in the data, attributes of the objects and relationships between them.
3. *Task complexity*: This is the load imposed by task demand including the number of objects involved and interactivity of these objects. This is an objective measure, and should be distinguished from task difficulty, which refers to how a task is experienced and perceived by the viewer, and can be affected by other factors, in addition to task complexity.
4. *Visual complexity*: Visual complexity includes all visual elements and their spatial distributions. It is determined by how these elements are visually represented, how well their spatial relationships match their intrinsic structural links, and to what extent the spatial arrangement conforms to human visual perception conventions and cognitive processes required by the task at hand. Therefore, a visualization with fewer elements or based on general aesthetics does not necessarily always lead to low visual complexity.
5. *Demographic complexity*: This includes motivation, age, gender, cognitive skills, domain knowledge and mental status. A large and growing body of literature has shown that these factors link to cognitive load in one way or another (for example, Wickens and Hollands[20]). For instance, the more domain knowledge the viewer has, the lower the effort will be needed in understanding biology networks.
6. *Time complexity*: Time complexity for visualization understanding can vary in different situations. It also varies depending on how the viewer perceives time availability. Time pressure imposed by reactions to different time conditions has 'the most profound and widespread effects on workload perceptions'.[29]

Note that the factors above affect cognitive load not only individually, but also interactively; this is what we call *between-factors complexity*. For example, a visualization that easily reminds the viewer of prior knowledge stored in memory may induce lower cognitive load.

*Assessment factors*

Assessment factors that can be affected by cognitive load include mental effort, response accuracy and response time. Each of them can be considered to reflect one aspect of cognitive sequences induced by cognitive
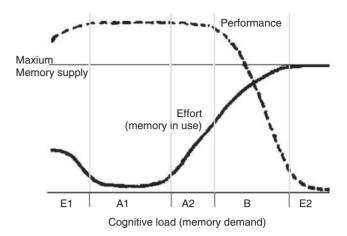


**Figure 3:** The model of task performance, mental effort and cognitive load.

load. In terms of visualization efficiency, response time and mental effort are classified as factors of cognitive cost, while response accuracy a factor of cognitive gain. Further, these three factors not only are affected by cognitive load, but also interact with each other. For example, spending more time may improve task accuracy. Increasing mental effort may reduce response time and increase accuracy.

**The model**

To further reveal how cognitive load and its assessment factors interact with each other under the limits of memory, we propose a theoretical model of task performance, mental effort and cognitive load, as outlined in Figure 3. The model was informed by the driver's workload model of De Waard[21] and the function of schematic resource supply and demand described by Wickens and Hollands.[20]

1. In Region E1, where memory demand is consistently low, performance may be negatively affected due to boredom or disregard. Extra effort may be also required to keep concentrated.
2. In Region E2, the demand on the viewer's memory exceeds resources available; the viewer is overburdened and performance is at a minimum level.
3. In Region A1, the viewer can comfortably handle the increase in memory demand, without having to devote extra effort; performance remains at a stable optimal level.
4. In Region A2, while demand increases, performance measures still remain unaffected. However, the viewer is only able to maintain the level of performance, by devoting more effort within the capacity of memory. With continuous increases in memory demand, exerting more effort is no longer workable to maintain performance optimal, a shift from Region A2 to Region B takes place.

5. In Region B, performance is affected and a decline appears. At a certain moment, memory demand will exceed the viewer's capability of effort compensation and Region E2 is entered, where performance has dropped to a minimum level.

From this model, we can see that keeping cognitive load low may not always lead to optimal performance. In many cases, low cognitive load causes boredom and concentration loss, resulting in the viewer disengaging with task performance. On the other hand, low cognitive load with complex data means that little information is being processed and effective visualization understanding is not happening.

Further, the proposed construct indicates that cognitive load can be induced by different factors. Current mental effort measures, however, reflect only overall cognitive load induced. Therefore, these measures are unable to tell how much of the cognitive load is caused by which factor. From a graph visualization point of view, cognitive load that is not related to the task at hand, such as cognitive load induced by visual complexity, should be reduced as much as possible. On the other hand, cognitive load that contributes to visualization understanding, such as cognitive load induced by data complexity, should be increased within the capacity of human memory.

It should be noted that when memory demand is high, it is true that the viewer may increase effort within the memory capacity to maintain the performance level. It is also possible that the viewer adopts an alternative 'more efficient, less source-consuming'[20] strategy to perform the task, with or without increased effort.

This region-based model shown in Figure 3 also demonstrates suitability of different measures in different situations. In region B, both performance and mental effort are good candidates for measuring cognitive load, while in Region A2, mental effort is more suitable. In Region A1, both mental effort and performance are insensitive to changes in memory demand. In this region, alternative measures other than performance and mental effort are required. Finally, in extreme situations such as those in Regions E1 and E2, performance and effort measures may suggest misleading information about the cognitive load. However, in the design stage, it may not be possible to know beforehand how much cognitive load is too much and in which region cognitive demand falls.

## User Study

Having introduced the cognitive load model and measures, we conduct a user study in the context of multi-relational social networks to test their validity. It is hoped that data patterns resulted from the experiment fit or help to refine our model, and demonstrate the usefulness of mental effort and visualization efficiency in providing insights that are not necessarily reflected in performance measures. Given the purpose of validation,

**Table 1:** Networks used in the experiment

| Network | 1 | 2 | 3 |
|---|---|---|---|
| Number of actors (size): | 16 | 20 | 25 |
| Number of links: | 36 | 59 | 98 |
| (Shared links included) | 10 | 9 | 12 |
| Links in relation 1 | 15 | 36 | 51 |
| Links in relation 2 | 21 | 23 | 27 |
| Links in relation 3 | | | 20 |

it is worth noting that in this study, it is not the context, but the resultant data that is more important. However, a careful design is needed to ensure the quality of the data.
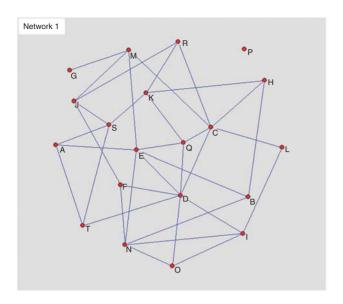
### Design

There are many ways available to visualize multi-relational networks. One way of them is to represent the network data in a single visualization using color to differentiate relations, which we call *Combined version*. Another way is to visualize the relation that is relevant to the task at hand only, with information from other relations faded in background or removed, which is called *Filtered version*. These visual versions were used as two conditions of visual complexity.

Three networks of different sizes were used (see Table 1). Before drawing the networks, it was decided that relation 1 was to be shown in the filtered version, and only information from this relation would be used for task performance to control the complexity of the study. Each of the three networks was drawn using the two visual versions. To avoid any possible bias that a layout is favorable to a particular task, for each version, the network was drawn twice. This produced two sets of diagrams (or *sociograms*) with each set having six diagrams (see Figure 4 for examples). These diagrams were produced using a force-directed method by Pajek,[30] which is a popular social network analysis and visualization tool among sociologists, with hand tuning to remove overlaps.

Graph reading is basically a visual search activity. Task complexity was determined based on the theoretical model proposed by Wood.[31] This model defines component complexity of a task as a function of the number of acts to be executed for the task and the number of information elements to be processed for performing those acts. Using search coverage or visual elements involved as a major consideration of component complexity, based on the analysis of cognitive processes of candidate tasks proposed beforehand, the following three levels of tasks were chosen:

1. One-Neighbor task (simple): Find one neighbor of a given actor.
2. Common-Neighbor task (medium): Find all common neighbors of two given actors.
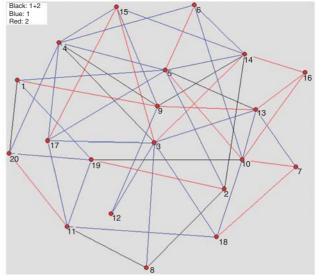3. Pattern task (complex): Find all triangle patterns.

**Figure 4:** Examples of stimuli with Filtered version on the left and Combined version on the right.

In determining the above tasks, to control possible confounding factors and to ensure the validity of the experiment, the following criteria had also been taken into consideration:

1. Tasks should not be too simple, and cannot be accomplished by simple perceptual operations.
2. Tasks should not be too complex in terms of task understanding and processing.
3. Task definitions themselves should be clear and short. Varying lengths of wording and numbers of components in task definition require different amounts of memory resources. In addition, a task that requires high effort to remember may compete for memory resources needed for task performance.
4. Tasks should be practical social network questions. But in-depth domain knowledge is not required.

For One-Neighbor and Common-Neighbor tasks, the given actors in each network were determined before the start of the experiment and were highlighted in green in the diagrams. The actors were chosen so that (1) for each of the two tasks, the number of links involved increased with the complexity of data; (2) for Common-Neighbor task, the two selected actors in each network had the same link distribution among the network relations. For example, in a two-relation network, if a chosen actor had two links for relation 1 and four links for relation 2 with one link shared between the two, then the other one should also have the same relation pattern. This means two links for relation 1 and four links for relation 2 with one link shared between the two. By doing this, it was ensured that no matter which actor to start with in looking for common neighbors, the same number of cognitive operations would be needed.

## Subjects

Subjects recruited were 30 students, who were enrolled with the School of Information Technologies of the University of Sydney. They were either postgraduate coursework or research students. All of them were familiar with node-link diagrams. When presented with a sociogram, all of them had no trouble performing tasks based on the social aspect of the underlying network, while none of them had formal academic or working experience related to social networks. The subjects participated in the experiment on a completely voluntary basis, and were reimbursed $20 each upon the completion of their assigned tasks.

## Procedure

The experiment was conducted in a computer laboratory room on an individual basis. A custom-built system was used to display the diagrams online. First the subject read through a tutorial sheet. Then the chance was given to ask questions and practice the system. The subject formally started the experiment by pressing the button on the screen.

For each subject, only one set of diagrams was randomly chosen for viewing. The diagrams were presented using a $6 \times 6$ Latin Square design to counteract order effects. Since there were three tasks to be performed, the same diagram was presented three times. The order of the tasks for each diagram was random. Each time a diagram was shown, the subject was required to answer the question as quickly as possible without compromising accuracy, and to press the button immediately after writing down the answer on the answer sheet provided. During viewing, it was not

**Table 2:** Means of response time, accuracy, mental effort and efficiency

|  | Visual complexity | | Data complexity | | | Task complexity | | |
|---|---|---|---|---|---|---|---|---|
|  | Filtered version | Combined version | Network 1 | Network 2 | Network 3 | One-neighbor | Common-neighbor | Pattern |
| Time (sec.) | 32.03 | 48.99 | 19.23 | 41.88 | 60.42 | 11.76 | 33.57 | 76.20 |
| Accuracy | 0.88 | 0.84 | 0.96 | 0.83 | 0.78 | 0.98 | 0.81 | 0.78 |
| Effort | 2.82 | 3.53 | 2.19 | 3.38 | 3.95 | 1.56 | 3.12 | 4.85 |
| Efficiency | 0.24 | −0.24 | 0.71 | −0.12 | −0.59 | 1.01 | 0.00 | −1.01 |

allowed to take notes, talk aloud or use a mouse to help. The response time was logged in real time starting when a diagram was completely shown and ending when the button was pressed. The subject continued on to indicate on the answer sheet the mental effort experienced based on the 9-point scale of Paas.[24] The subject pressed the button again to proceed to the next diagram. After the online tasks, the subject finished the experiment with a simple questionnaire. The whole session took about 50 minutes.

## Results

A within-subjects 2 (visual complexity) ×3 (data complexity) ×3 (task complexity) design was employed in the experiment. It was expected that each of the independent variables: visual complexity, data complexity and task complexity had a significant effect on the dependant variables: time, accuracy, mental effort and visualization efficiency. Efficiency was computed using Equation (4). In regard to accuracy, it was recorded as 1 if a given answer was correct, and 0 otherwise for One-Neighbor and Common-Neighbor tasks. For Pattern task, accuracy was measured as how far the response was from zero, and calculated in the following way: if a given answer was no more than the correct answer, accuracy was recorded by dividing the given answer by the corresponding correct answer. Otherwise, accuracy was obtained by subtracting the difference between the given answer and the corresponding correct answer from the correct answer first, then dividing the obtained value by the correct answer; if the obtained value is less than 0, then 0 was recorded. Therefore the accuracy values range from 0 to 1; the higher the accuracy value, the better the performance. This way of measuring helps to maintain accuracy within the same range as for the three tasks, thus avoiding unbalanced weights in accuracy when these three tasks are analyzed as a whole.

Table 2 contains the means of response time, accuracy, mental effort and efficiency. The statistical results of the analysis of variance (ANOVA) with repeated measures are shown in Table 3. In this table, along with $F$ statistic and $P$ values, $\eta^2$, a measure of effect size, is reported as well. $\eta^2$ indicates how large an effect of one variable on another is, while $P$ tells us whether this effect is statistically significant or not. According to common rules of

thumb, $\eta^2$ is small at 0.01, medium at 0.10 and large at 0.25.

Considering the visual complexity effect, it can be seen from Table 2 that the subjects spent more time and experienced higher effort with Combined version yielding lower response accuracy, compared to Filtered version. The efficiency measure indicated that Combined version was less efficient. ANOVA with repeated measures (see Table 3) revealed that visual complexity had a significant effect on response time, on mental effort and on efficiency, but not on accuracy.

In regard to the data complexity effect, Table 2 shows that on average, response time and mental effort increased with data complexity, while accuracy and efficiency decreased. ANOVA with repeated measures (see Table 3) revealed that data complexity had a significant effect on response time, on accuracy, on mental effort and on efficiency.

The task complexity effect followed a similar pattern. As can be seen from Table 2, response time and mental effort increased with task complexity, while accuracy and efficiency decreased. ANOVA with repeated measures (see Table 3) revealed that task complexity had a significant effect on response time, on accuracy, on mental effort and on efficiency.

In regard to the between-factors complexity effect, ANOVA with repeated measures (see Table 3) revealed that each pair of the complexity factors interactively affected the dependant variables with one exception. There was no interaction effect between visual complexity and data complexity on accuracy.

## Discussion

The experimental results reinforce the conceptual construct of cognitive load proposed; cognitive load can be induced by various complexity factors, and the induced cognitive load is reflected not only in the performance measures but also in the mental effort measure. Due to the effective manipulation of the experimental conditions, the data patterns follow our expectations. This demonstrates that it is workable to use the 9-point scale of Paas[24] in measuring mental effort and to use Equation (4) in measuring visualization efficiency.

The usefulness of mental effort in revealing detailed cognitive cost for visualization cognition can be seen

**Table 3:** ANOVA with repeated measures on response time, accuracy, mental effort and efficiency

| Source | df | Time | | Accuracy | | Effort | | Efficiency | |
|---|---|---|---|---|---|---|---|---|---|
| | | F | $\eta^2$ | F | $\eta^2$ | F | $\eta^2$ | F | $\eta^2$ |
| Visual complexity(V) | 1 | 79.36*** | 0.73 | 3.91 | 0.12 | 99.91*** | 0.78 | 77.20*** | 0.73 |
| Error (V) | 29 | (489.27) | | (0.07) | | (0.68) | | (0.39) | |
| Data complexity (D) | 2 | 77.34*** | 0.73 | 22.05*** | 0.43 | 94.28*** | 0.77 | 202.47*** | 0.88 |
| Error (D) | 58 | (990.50) | | (0.07) | | (1.53) | | (0.39) | |
| Taskcomplexity (T) | 2 | 64.85*** | 0.69 | 22.53*** | 0.44 | 144.60*** | 0.83 | 183.62*** | 0.86 |
| Error (T) | 58 | (2981.24) | | (0.09) | | (3.37) | | (0.99) | |
| V × D | 2 | 12.73*** | 0.31 | 0.25 | 0.01 | 4.60* | 0.14 | 5.80** | 0.17 |
| Error (V × D) | 58 | (336.63) | | (0.05) | | (0.60) | | (0.29) | |
| V × T | 2 | 15.47*** | 0.35 | 4.64* | 0.14 | 9.93*** | 0.26 | 16.27*** | 0.36 |
| Error (V × T) | 58 | (502.14) | | (0.08) | | (0.57) | | (0.37) | |
| D × T | 4 | 40.91*** | 0.59 | 7.65*** | 0.21 | 56.98*** | 0.66 | 69.84*** | 0.71 |
| Error (D × T) | 116 | (841.02) | | (0.06) | | (0.94) | | (0.40) | |
| V × D × T | 4 | 3.36* | 0.10 | 1.98 | 0.06 | 2.92* | 0.09 | 1.88 | 0.06 |
| Error (V × D × T) | 116 | (317.79) | | (0.04) | | (0.47) | | (0.25) | |

*Note*: Numbers enclosed in parentheses represent mean square errors.
*$P < .05$; **$P < .01$; ***$P < .001$.

**Table 4:** Means of response time, accuracy, mental effort and efficiency and ANOVA with repeated measures for network 1 and One-Neighbor task.

| | Filtered version | Combined version | F(1,29) | $\eta^2$ |
|---|---|---|---|---|
| | Mean | Mean | | |
| Time | 10.77 | 11.39 | 0.21 | 0.01 |
| Accuracy | 1.00 | 0.97 | 1.00 | 0.03 |
| Effort | 1.40 | 1.67 | 6.27* | 0.18 |
| Efficiency | 1.09 | 0.95 | 3.16 | 0.10 |

*$P < .05$.

from the effects of visual complexity (see Tables 2 and 3). The subjects achieved similar levels of accuracy with an average at 0.84 for the combined version and 0.88 for the filtered version ($F(1, 29) = 3.91$, $P > 0.05$, $\eta^2 = 0.12$). However, analysis indicated that to maintain accuracy at the similar levels, the subjects had exerted significantly more effort with the combined version than with the other and this difference is not trivial in terms of effect size ($F(1, 29) = 99.91$, $P < 0.001$, $\eta^2 = 0.78$). This situation is reflected in Region A2 of the model in Figure 3. It supports one of our arguments that new measures, in addition to traditional performance measures, should be employed to be able to evaluate graph visualizations accurately.

A further example can be seen from the experimental data of the two visual versions when One-Neighbor task and network 1 were used. The means of individual measures for the two versions and the results of ANOVA with repeated measures are provided in Table 4. Overall, the magnitude of the visual complexity effect is small for time and accuracy and medium for effort and efficiency in this specific case. Neither response time nor accuracy was significantly affected by visual complexity. However, the difference in visual version was detected by the mental effort measure; the subjects devoted significantly higher effort with the combined version than they did with the filtered version ($F(1, 29) = 6.27$, $P < 0.05$, $\eta^2 = 0.18$). If only performance was measured, we would have concluded that the two visual versions were equivalent.

Considering $P$ and $\eta^2$ values from Tables 3 and 4, it appears that the mental effort measure was slightly more sensitive to changes in the experimental conditions, compared to performance measures. This is reasonable since mental effort was measured as actual cognitive load allocated for task performance. Changes in conditions, therefore in actual cognitive load, should have immediate reflection in mental effort. On the other hand, performance may not have such an immediate reflection as mediating factors may act in between offsetting those changes. For instance, the viewer may adopt an alternative strategy to process information when cognitive load changes. This good sensitivity is in agreement with prior finding on the rating scale of Paas,[26] and is also demonstrated by the proposed model illustrated in Figure 3. From this figure, it can be seen that mental effort is sensitive to a wider range of cognitive demand on memory. With the increase in memory demand, mental effort increases across Regions A2 and B, while performance remains unchanged in Region A2.

Despite the advantages of mental effort in achieving better delineation of cognitive costs, individual measures may not be so useful. There are situations where such measures may not be consistently in favor of one visualization against another. In addition, the measured cognitive load can only be meaningfully interpreted in association with the corresponding performance level. The proposed visualization efficiency overcomes these issues by

considering all the measures at the same time. Visualization efficiency makes it possible to assess the overall quality of different visualizations comparatively. In the context given by Equation (4), visualization efficiency assesses the attained performance level (accuracy) relative to the effort devoted and time spent during task performance.

Further, in Table 4, visualization efficiency appears somehow to have smoothed out the effects of visualization complexity on the performance and mental effort measures. A significant difference in a measure may not necessarily lead to a significant difference in efficiency, and the values of effect size for efficiency were roughly in the middle range among those for all measures. This may be argued as an additional benefit of using the efficiency measure. Subjects often have to make trade-offs throughout task performance. For example, subjects may either spend more time and make more accurate answers, or *vice versa*; they may also either invest more effort and spend less time, or *vice versa*. Random fluctuations resulting from these kinds of tradeoffs in the individual measures are smoothed out when these measures are incorporated into a single measure. In this sense, the combined efficiency measure (Equation (4)) not only enables us to evaluate visualizations at the overall level, but also might be more reliable than other measures.

The results of this experiment have shown that when a graph becomes relatively large and dense, even with only 25 nodes and 98 links (network 3), human perception and cognitive systems can quickly be overburdened causing errors in performing relatively complex tasks, for example, finding relationship patterns (Pattern task). As revealed by analysis of cognitive processes, indeed, to find all triangles of relation 1 from network 3, the combined version resulted in much higher element interactivity,[15] compared to the filtered version. That is, the subjects had to simultaneously process a larger number of interacting elements held in working memory, as follows:

1. Go to the legend area, find and remember which color represents relation 1;
2. Remember the number of triangles that have been found so far;
3. Remember the triangles that have been found, so that they will not be counted again;
4. Remember the number of links that have so far been searched to form a new triangle;
5. Remember the links that have been searched to form a new triangle, so that they will not be counted again;
6. Each time, when searching for a new link, compare the color of the link with the one held in memory to decide whether it belongs to relation 1;
7. Once a new link is found, add 1 to the number of links having been found and hold the number in memory;
8. Once a new triangle is found, add 1 to the number of triangles having been found and hold the number in memory;
9. If the relation 1 color is forgotten, go back to the legend area again to find it.
10. If the triangles or links or the number of links having been searched are forgotten, go back to the previous step and search again or restart the whole task process;
11. If the number of triangles having been found is forgotten, restart the whole task process.

This cognitive process can become particularly effortful and time-consuming when the number of nodes, links, relations and triangles increases. The comments of the subjects showed that high cognitive load may have emotional implications and result in changes of task performing strategies. Their reactions to high cognitive load also differed from subject to subject. For example, 'It might involve some guessing'; 'Start searching from easy parts, then move to hard parts,' 'Count triangles randomly, not from node to node'; 'Change from selecting node randomly to search from node by node'; 'I'll try my best for a while, if no hope, then give up'; 'If one way doesn't work, I will try another way to find answers'; 'So easy to lose track, I have to try again and again'; 'Be more careful'; 'Slow down'; 'Double check'; 'A bit nervous'; 'Felt dazzling'; 'Calm down' and so on.

When visualizing large and complex data, the efficient use of limited human cognitive capacity stands out and becomes an important factor determining the success of a particular visualization technique. When tasks are complex, performance may not be automatically improved by adopting better visualizations; capacity of human perception and cognition also has a role to play.[3] It should be noted that apart from those factors affecting cognitive load, visualization effectiveness can be also affected by many other factors.[32] However, since any visual information has to pass through the human sensory system first, and to be processed in working memory, failure to respect human cognitive capacity might undo any positive benefits of new technologies and techniques introduced for visualization.

It may be argued that it does not matter how much mental effort is devoted as long as it is within the cognitive capacity. On the one hand, it may be true when tasks are easy and data sets are simple. As can be seen from the model showed in Figure 3, optimal performance can still be achieved by increasing effort when the visualization is not the best. On the other hand, with the rapid increase in size and complexity of data, visualization understanding has become a prolonged and cognitively intensive process. Persistent high effort may cause, for example, fatigue, which in turn leads to further increase of effort. In addition, increased mental effort unnecessarily consumes more memory resources, thus competing with simultaneous decision making processes for the memory resources needed.

It may be also argued that any difference in mental effort devoted may not be that important as long as the optimal level of performance can be reached. However,
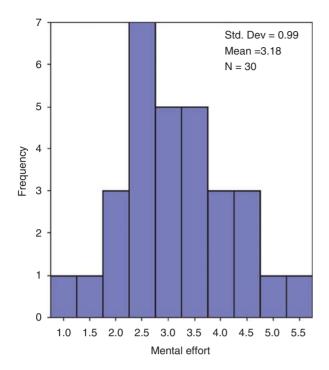
**Figure 5:** Distribution of mean mental efforts.

from a designer's point of view, it is helpful to examine why the two different visualizations induce different amounts of cognitive load. Understanding this certainly helps to devise better visualizations and make more accurate assessments.

It is worth mentioning that although it is tempting to conclude that the filtered version is more efficient than the combined version from this particular study, and indeed the filtered version is visually less complex with only relevant information being displayed, caution should be taken in generalization. The experiment was not intended to systematically exam the strengths and weakness of the two visual versions; the three tasks used were relevant to only one of the relations, which was consistently in favor of the filtered version. There are occasions when the combined version or other types of visualizations are more desirable.[33]

*Demographic complexity and time complexity*
In this experiment, one important factor that was not included in our investigation is demographical complexity. Figure 5 shows the distribution of means of mental effort devoted by each subject. It can be clearly seen that the devoted mental effort differed greatly across the subjects. Analysis of variance also indicated that this difference across the subjects was statistically significant ($F(29, 510) = 4.31$, $P < 0.001$, $\eta^2 = 0.19$). In addition, plenty of empirical evidence is also available in the literature suggesting that cognitive capacity and the level of ability to process information differ with demographical

factors, such as gender, age, cognitive style, cognitive skill, prior knowledge and spatial ability.[15,20] These differences induce different levels of cognitive load and affect performance to different extents. However, specific research in the domain of graph visualization is needed to investigate how those human factors interact with different visualizations.[1]

As regard to time complexity, during the test, the time requirement was set by asking subjects to perform tasks as quickly as possible without compromising accuracy; time was a dependant variable, instead of a fixed time limit. Research in psychology, cognitive science and application domains has suggested that there is an impact of time availability on perceived mental effort and task difficulty (for example, Kerick and Allender[29]). Increasing time pressure induces anxiety[34] and increased anxiety leads to task-irrelevant thoughts.[35] These thoughts consume working memory resources and may impair cognitive task performance. In addition, during tests, it is possible that subjects interpreted the time requirement differently and restricted the time availability at different levels, and changes in time availability may affect their cognitive strategies and behaviors in performing tasks, therefore bringing possible confounding noises into the resulting data.

## General Discussion

### Visualization efficiency

The three-dimensional method of measuring visualization efficiency (see Equation (4)) can also be adapted according to different analysis needs or variables available. For example, studies evaluating graph visualizations usually measure user preference (PRE) instead of mental effort (ME), in addition to response accuracy (RA), and response time (RT). In this case, Equation (4) can be modified accordingly by replacing ME with PRE:

$$E = \frac{z_{RA} + z_{PRE} - z_{RT}}{\sqrt{3}} \qquad (6)$$

In the context given by Equation (6), high visualization efficiency is achieved when high accuracy is attained in association with high user preference and a short response time. Equation (6) treats every variable equally. Alternatively, we might assign different weights ($w_i$, where $i = 1, 2, 3$) to these variables according to individual priorities in the overall evaluation. Therefore Equation (6) can be refined as follows:

$$E = \frac{w_1 z_{RA} + w_2 z_{PRE} - w_3 z_{RT}}{\sqrt{3}} \qquad (7)$$

where

$$\sum_{i=1}^{3} w_i = 1.$$

Further, in cases where only RA and RT are measured, a similar two-dimensional approach to that of Paas and

Van Merrienboer[27] can be used to compute visualization efficiency as follows:

$$E = \frac{z_{RA} - z_{RT}}{\sqrt{2}} \tag{8}$$

To be more general, given $m$ cognitive cost measures $X_i$ ($i = 1, 2, 3, \ldots, m$) and $n$ cognitive gain measures $Y_j$ ($j = 1, 2, 3, \ldots, n$), we can simply group them into a combined cognitive cost measure $\sum_{i=1}^{m} X_i$ and a combined cognitive gain measure $\sum_{j=1}^{n} Y_j$, and determine E as follows:

$$E = \frac{\sum_{j=1}^{n} z_{Y_j} - \sum_{i=1}^{m} z_{X_i}}{\sqrt{2}} \tag{9}$$

It is important to note that relying solely on the visualization efficiency measure can sometimes be misleading. Findings based on efficiency are not always identical to those based on effectiveness measures.[14] In other words, visualization efficiency should be treated as a supplementary rather than an alternative measure of traditional effectiveness measures. After all, performance is such an important criterion for visualization evaluation. Evaluating visualizations in terms of not only their relative efficiency, but also their effectiveness, allows us to make more comprehensive and more accurate judgments about their quality.

Another limitation of the efficiency measure is that the proposed equations assume a linear relationship between the measures, as was the case in the original instructional efficiency calculations of Tuovinen and Paas[26] and Paas and Van Merrienboer. [27] Although a number of experiments conducted in cognitive load research have demonstrated the added value of the efficiency measure,[14] this assumption is clearly an oversimplification and the reality can be more complex, as shown in Figure 3.

### Related work and cognitive load considerations

There has been a growing interest in cognitive load considerations in the visualization literature. Fischer *et al*[36] measured cognitive load using NASA TLX in investigating how animation speed affects instructional effectiveness of dynamic visualizations. In a study concerning the effects of large displays on human visualization perception and cognition, Yost and North[37] also measured work load using a modified version of NASA TLX. Purchase *et al*[38] conducted a study examining how important the mental map is for understanding dynamic graphs; it was found that mental map matters for node identification tasks. Ware *et al*[9] introduced a methodology that allows us to investigate cognitive costs of multiple graph features in the same experiment. Jankun-Kelly[39] introduced visualization process graphs and presented methods of visualizing those graphs for understanding how visualizations are explored. Henry *et al*[40] described a method of automatic logging and replay mechanisms for longitudinal

studies. Shrinivasan and Van Wijk[41] proposed a three-view visualization framework for supporting analytical reasoning and demonstrated its usefulness in a user study. Ziemkiewicz and Kosara[42] presented an approach of evaluating visualizations in terms of visual metaphors, representing a significant step toward a better understanding of how different visualizations affect cognitive processes.

Despite the studies mentioned above, cognitive load considerations with visualizations require further attention from the research community. The implications of cognitive load considerations are twofold. First, apart from the benefits of using cognitive measures discussed in previous sections, cognitive load considerations also make more realistic evaluations possible. While in prior research, simple tasks have been used for evaluation, complex tasks are not often used. Absence of complex tasks has been a great threat to generalization of evaluation findings as real tasks are usually complex. In general, complex tasks require intensive and prolonged information processing in memory, which may overburden the viewer causing interference of confounding factors. Understanding cognitive processes helps in identifying those confounding factors. And at the same time, knowing how information is to be processed is an important part of visualization evaluation.[42] While performance measures tell us the end results of a cognitive process, mental effort and visualization efficiency indicates what is happening *during* the process. Indeed, mental effort and efficiency have been used as major measurements in research of CLT for more than a decade. Empirical studies on CLT are typically conducted in learning environments, in which tasks are usually difficult and complex.

Second, cognitive load considerations have implications on visualization design. One of the key features of effective visualizations is to enable people to gain insights into information that are not immediately apparent in the visualization.[43,44] Cognitive reasoning is one of the main ways to obtain this kind of insight; visualizations are more beneficial if they also support human reasoning.[2] Information retrieving and reasoning require sufficient memory resources to be successful.[12] One way of facilitating these processes is to minimize cognitive load induced by visual complexity. By doing so, more memory resources can be allocated for cognitive load imposed by data complexity and task complexity. These kinds of load should be increased so that more information can be visualized and effectively conveyed to the viewer. In other words, minimizing visual cognitive load and maintaining overall load at a *reasonable* level at the same time, not only makes the most of the limited human cognitive capacity, but also fosters maximum information retrieving and effective processing.

### Conclusion

Evaluation has become increasingly important and crucial in ensuring effectiveness.[13,45,46] Traditional performance

measures have some limitations in detecting differences between visualizations. Therefore they are not always sufficient in measuring effectiveness. Inconsistency between individual measures in supporting a visualization makes it difficult to judge its overall quality and provides little practical guidance in visualization design. This paper addresses these issues from a cognitive load perspective, based on recent results of cognitive load research.

Cognitive load and visualization efficiency measures are introduced to overcome the limitations of performance-based measures. A conceptual construct of cognitive load and a theoretic model of task performance, mental effort and cognitive load are proposed to reveal their interacting relationships during visualization understanding. The experiment presented in this paper demonstrates the validity and usefulness of the proposed model and measures. Different equations for the efficiency measure are also suggested to meet specific evaluation needs. Clearly, validity and reliability of the proposed model and measures in the domain of graph visualization still need further examinations.

## Acknowledgements

## References

1 Tory, M. and Moller, T. (2004) Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* 10(1): 72–84.
2 Card, S., Mackinlay, J. and Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers.
3 Lohse, G.L. (1997) The role of working memory on graphical information processing. *Behavior & Information Technology* 16(6): 297–308.
4 Baddeley, A. (1986) *Working Memory*. Oxford, UK: Oxford University Press.
5 Zhang, J., Johnson, K.A., Malin, J. and Smith, J.W. (2002). Human-centered information visualization. International Workshop on Dynamic Visualization and Learning; 18–19 July, Tubingen, Germany.
6 Purchase, H. (1998) Performance of layout algorithms: Comprehension, not computation. *Journal of Visual Languages and Computing* 9: 647–657.
7 Purchase, H. (2005). Graph drawing empirical studies: Lessons learned. Dagstuhl Seminar 05191 on Graph Drawing: Wadern, Germany.
8 Robertson, G. (2000). Keynote lecture: Leveraging human capabilities in information perceptualization. Proceedings of the Fourth International Conference on Information Visualisation (IV' 00). Los Alamitos, CA: IEEE Computer Society, pp. 3.
9 Ware, C., Purchase, H., Colpoys, L. and McGill, M. (2002) Cognitive measurements of graph aesthetics. *Information Visualization* 1(2): 103–110.
10 Herman, I., Melancon, G. and Marshall, M. (2000) Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* 6(1): 24–43.
11 Ghoniem, M., Fekete, J.-D. and Castagliola, P. (2004). A Comparison of the readability of graphs using node-link and matrix-based representations. Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis' 04); 10–12 October, Washington, DC: IEEE Computer Society, pp. 17–24.
12 Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. In: S.P. Robertson, G.M. Olson and J.S. Olson (Eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology (CHI' 91); 27 April—2 May, New Orleans, Louisiana, New York: ACM, pp. 137–144.
13 Plaisant, C. (2004). The challenge of information visualization evaluation. Proceedings of the Working Conference on Advanced Visual Interfaces (AVI' 04), 25–28 May, Gallipoli, Italy. New York: ACM, pp. 109–116.
14 Paas, F., Tuovinen, J., Tabbers, H. and van Gerven, P. (2003) Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist* 38(1): 63–71.
15 Van Merrienboer, J. and Sweller, J. (2005) Cognitive load theory and complex learning: Recent development and future directions. *Educational Psychology Review* 17(2): 147–177.
16 Atkinson, R.C. and Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In: K.W. Spence and J.T. Spence (eds.) *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 2. New York: Academic Press.
17 Miller, G. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63(2): 81–97.
18 Peterson, L. and Peterson, M. (1959) Short-term retention of individual verbal items. *Journal of Experimental Psychology* 72: 221–231.
19 Paas, F., Tuovinen, J., Van Merrienboer, J. and Darabi, A. (2005) A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development* 53(3): 25–34.
20 Wickens, C. and Hollands, J. (1999) *Engineering Psychology and Human Performance*. 3rd edn. Upper Saddle River, NJ: Prentice-Hall.
21 De Waard, D. (1996). The measurement of drivers' mental workload. PhD thesis, University of Groningen, Haren, The Netherlands.
22 Gopher, D. and Braune, R. (1984) On the psychophysics of workload: Why bother with subjective measures? *Human Factors* 26: 519–532.
23 Hendy, K.C., Hamilton, K.M. and Landry, L.N. (1993) Measuring subjective workload: When is a one scale better than many? *Human Factors* 35(4): 579–601.
24 Paas, F. (1992) Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology* 84: 429–434.
25 Marcus, N., Cooper, M. and Sweller, J. (1996) Understanding instructions. *Journal of Educational Psychology* 88(1): 49–63.
26 Tuovinen, J. and Paas, F. (2004) Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science* 32: 133–152.
27 Paas, F. and Van Merrienboer, J. (1993) The efficiency of instructional condition: An approach to combine mental effort and performance measures. *Human Factors* 35: 734–743.
28 Paas, F. and Van Merrienboer, J. (1994) Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review* 6(4): 351–371.
29 Kerick, S., and Allender, L. (2004). Effects of cognitive workload on decision accuracy, shooting performance, and cortical activity of soldiers, ⟨http://handle.dtic.mil/100.2/ADA433487⟩.
30 Batagelj, V. and Mrvar, A. (2003). Pajek: Analysis and visualization of large networks. In: M. Junger and P. Mutzel (eds.) *Graph Drawing Software*, Berlin, Heidelberg: Springer-Verlag, pp. 77–103.
31 Wood, R. (1986) Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37(1): 60–82.

32 Kalyuga, S. (2000) When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology* 16(2): 161–172.

33 Yost, B., and North, C. (2005). Single complex glyphs versus multiple simple glyphs. Proceedings of (CHI' 05) Extended Abstracts on Human Factors in Computing Systems (CHI' 05); 2–7 April, Portland, OR: New York: ACM, pp. 1889–1892.

34 Tohill, J. and Holyoak, K. (2000) The impact of anxiety on analogical reasoning. *Thinking and Reasoning* 6: 27–40.

35 Eysenck, M. (1979) Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality* 13: 363–385.

36 Fischer, S., Lowe, R. and Schwan, S. (2008) Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology* 22: 1126–1141.

37 Yost, B. and North, C. (2006) The Perceptual scalability of visualization. *IEEE Transactions on Visualization and Computer Graphics* 12(5): 837–844.

38 Purchase, H., Hoggan, E. and Gorg, C. (2007). How Important is the 'Mental Map'? – An Empirical Investigation of a Dynamic Graph Layout Algorithm. Proceedings of 14th International Symposium on Graph Drawing (GD' 06), Lecture Notes in Computer Science, Vol. 4372, Berlin, Heidelberg: Springer-Verlag, pp. 184–195.

39 Jankun-Kelly, T.J. (2008). Using visualization process graphs to improve visualization exploration. In: J. Freire, D. Koop and L. Moreau (eds.) Provenance and Annotation of Data and Processes: Second International Provenance and Annotation Workshop (IPAW 2008); 17–18 June, Salt Lake City, UT. Revised Selected Papers. Lecture Notes in Computer Science, Vol. 5272. Berlin, Heidelberg: Springer-Verlag, pp. 78–91.

40 Henry, N., Elmqvist, N. and Fekete, J.-D. (2008). A methodological note on setting up logging and replay mechanisms in infovis systems. Position paper presented in BEyond time and errors: Novel evaLuation methods for Information Visualization (BELIV'08); April 2008.

41 Shrinivasan, Y. and van Wijk, J. (2008). Supporting the analytical reasoning process in information visualization. Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems (CHI' 08); 5–10 April, Florence, Italy. New York: ACM, pp. 1237–1246.

42 Ziemkiewicz, C. and Kosara, R. (2008) The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics* 14(6): 1269–1276.

43 Amar, R. and Stasko, J. (2005) Knowledge Precepts for Design and Evaluation of Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11(4): 432–442.

44 North, C. (2006) Visualization viewpoints: Toward measuring visualization insight. *IEEE Computer Graphics & Applications* 26(3): 6–9.

45 Chen, C. and Czerwinski, M. (2000) Empirical evaluation of information visualization: An introduction. *International Journal of Human-computer Studies* 53(5): 631–635.

46 Kerren, A., Ebert, A., and Meyer, J. (eds.) (2006). Human-Centered Visualization Environments. Dagstuhl Castle, Germany: GI-Dagstuhl Research Seminar, LNCS 4417, Springer-Verlag.