An exploratory study of cognitive load in diagnosing patient conditions

MICHAEL WORKMAN1, MICHAEL F. LESSER2 AND JOONMIN KIM3

1College of Business, Florida Institute of Technology, 150 West University Blvd, Melbourne, FL 32901, USA, 2LMI of Brevard, Inc. Holmes Regional Medical Center, 1350 S Hickory Street, Melbourne, FL 32901, USA, and 3College of Information, Florida State University, 101 Louis Shores Building, Tallahassee, FL 32306, USA

Abstract

Objective. To determine whether the ways in which information is presented to physicians will improve their ability to respond in a timely and accurate manner to acute care needs. The forms of the presentation compared traditional textual, chart and graph representations with equivalent symbolic language representations. To test this objective, our investigation involved two studies of interpreting patient conditions using two forms of information representation. The first assessed the level of cognitive effort (the outcome variable is known as cognitive load), and the second assessed the time and accuracy outcome variables.

Participants. Our investigation consisted of two studies, the first study involved 3rd and 4th year medical students, and the second study involved three board certified physicians who worked in an intensive care unit of a metropolitan hospital.

Design. The first study utilized an all-within-subject design with repeated measures, where pretests were utilized as control covariate for prior learning and individual differences. The second study utilized a random sampling of records analyzed by two physicians and qualitatively evaluated by board-certified intensivists.

Results. The first study indicated that the cognitive load to interpret the symbolic representation was less than those presented in the more traditional textual, chart and graphic form. The second study suggests that experienced physicians may react in a more timely fashion with at least the same accuracy when the symbolic language was used than with traditional charts and graphs.

Conclusions. The ways in which information is presented to physicians may affect the quality of acute care, such as in intensive, critical and emergency care units. When information can be presented in symbolic form, it may be cognitively processed more efficiently than when it is presented in the usual textual and chart form, potentially lowering errors in diagnosis and increasing the responsiveness to patient conditions.

Keywords: acute care, cognitive load, visual information representation

‘Information overload’ has become a common term in knowledge-work, where people must consider the relationships among many variables (integrated tasks) as well as the values of individual variables (focused tasks) [1–3]. Physicians often face this in the delivery of emergency care when they evaluate relationships among indicators of illnesses. Although tremendous importance has been placed on physician education with respect to physical examination, differential diagnosis and the performance and analysis of specialized tests, which might be termed ‘world knowledge,’ little emphasis has been placed on improving how information is presented, or what might be termed ‘display knowledge,’ which the physician needs to make timely and accurate decisions [2, 4, 5].

Despite using sophisticated medical diagnostic and therapeutic technology, most modern intensive care units continue to use the standard patient chart and flow sheet and a bedside monitor showing three or four wave formations for information display, and as computerized systems have been introduced into the intensive care setting, rather than redesigning the system to take advantage of their power, the computer has simply been used as a more convenient filing cabinet and the computer flow sheet appears little different from its paper cousin. Yet an important distinction has been made recently between design for ‘data availability’ and

Address reprint requests to: Michael Workman, College of Business, Florida Institute of Technology, 150 West University Blvd, Melbourne, FL 32901, USA. E-mail: workmanfit@yahoo.com/workmanfsu@yahoo.com
design for ‘information extraction,’ which pertain directly to the nature of information representations [6]. Information displays that consider data availability alone (such as the patient chart) often leave the physician with the burden of collecting relevant data, maintaining the data in working memory and mentally integrating these data to arrive at a decision. This process taxes available cognitive resources (the tax is called cognitive load [7, 8]). The distinction between availability and extraction is important because performance on a task is inversely related to cognitive load [9]. When cognitive load increases, there is generally a deterioration of performance observed in lower response times and greater errors in task performance.

To help determine whether information might be represented to physicians differently with a focus on data extraction in ways that would reduce cognitive load, we conducted two studies. The first study used a dual-task test [7, 10, 11] to isolate the cognitive load in making determinations about patient conditions using the traditional textual, chart and graph representations compared with equivalent symbolic language representations. In the second study, we were interested in the performance implications of a reduction in cognitive load. For this, the two forms of information were compared on time and accuracy in making assessments of patient conditions.

**Theory and background**

**Human–computer interaction and knowledge representation**

The standard patient chart and flow sheet represent information in textual form and a linear fashion [3–5]; they show a series of data point relationships over time, such as heart rate. As more information is presented in this form, people have increased difficulty in extracting the important information from the available data, because the complexity of information is partly due to the linear form in which the information is rendered [10, 11]. However, as pointed out by Langer [12], some forms of information may be more simultaneously apprehended and perceived more holistically than others, and that these more simultaneous forms of information representation should be perceived with less cognitive load than those in a linear form [3].

For instance, when examining artwork, people take in the image more holistically than when reading prose, and when looking at an image on a computer screen, people do not read the picture pixel by pixel [3, 5]. It is important to note that when people are asked to describe the meaning of such artwork or pictures, they are less capable of describing it in objective terms [13]; they are left with vague impressions and subjective interpretations of the information because there are no commonly accepted rules or vocabulary to describe what the artist intended to convey and what observer apprehended.

The rules or vocabulary for conveying information determine the objective measures by which people generally draw conclusions and make inferences about an intended meaning. To illustrate, the meaning of the word ‘tear’ is unclear unless predicated that, you have a tear in your shirt versus you have a tear in your eye. The basis for these rules is Chomsky’s [14] transformation grammar in which he pointed out that sentences exist on two levels, a deep and a surface structure. The deep structure is essentially that of meaning encoded with the surface structure, which is that of syntax. Transformational grammar is thus the system of rules that transforms a sentence from one structural level to another [15]. Therefore to formulate a conception of meaning, the relationships among the words or concepts must be known.

**Knowledge representation and cognitive load**

Studies of computer visualization technologies [16] suggest at least anecdotally that using symbols (called glyphs) that express a pattern or syntax to represent a certain meaning may reduce cognitive load to comprehend compared to traditional linear information representations, because they present information more holistically (called simultaneity) [17, 18]. Glyphs are pictures or figures that imply a concept that can be universally understood regardless of spoken language, such as the sign for ‘do not enter’ may consist of a red circle with a white horizontal line through it. Thus glyphs share several features with the simultaneous representational forms described by Langer [12]. In particular, they can assert symbolic meaning and because they are discrete forms, they are visually apprehended more or less at once as suggested in studies of comprehension and visualization [16] and examinations of eye saccades in scanning prose versus glyphs [5, 19, 20].

The reason why simultaneity affects cognitive load is because working memory is limited in its capacity, which is increasingly impinged as information becomes increasingly complex [21], and linearity is a major factor in information complexity [22–25]. The resulting deterioration in task performance often appears as a gradual decline rather than a calamitous breakdown, but the decline is measurable, and performance crucially depends on the relationship between working memory and cognitive load [10]. Bourke and Duncan [11] found using a visual search task that when the complexity of information increased, there was increasing, cognitive interference even among dissimilar concurrent tasks that led to an increase in task errors.

Underpinning this is a theory of implicit and explicit cognition [18]. Implicit cognition results from automatic processes, which are effortless, unconscious and involuntary [21], and explicit cognition results from intentional cognitive processes, which are effortful and conscious [17]. The intentional setting of the goals and intentional evaluation of outputs are called cognitive monitoring. Thus, a cognitive process is implicit if it has acquired the ability to run without cognitive monitoring, whereas intentional cognition requires cognitive monitoring and relies on working memory.

Although the retrieval of words is automatic and serves to cognitively prime meaning [17], the meaning construction and inference from are partly intentional integrative processes that
tax working memory, and this can be measured by dividing cognitive processes using the so-called dual-task tests [10, 15, 17, 18] in which two isolated stimuli are presented with variable stimulus onset asynchronies, and the reaction times are recorded. The result is that the reaction time to the secondary response is delayed with cognitive load, because the primary task utilizes resources needed to initiate the second task [19].

Considering the potential that numerical data may be more efficiently assimilated when presented graphically [13], numerous attempts have been made to display various types of medical information with various graphical formats, as clearly not all of the possible formats for portraying information result in the same effectiveness (Fig. 1) [2]. Examples are the standard line graph of time versus temperature common on a standard ‘flow sheet,’ the more elaborate ‘face’ display of Chernoff [26] (see also [27]), or the star-like Kiviat displays [28–30].

Although the line graph is well accepted, it is quite limited and almost useless for displaying a large number of variables. Although Chernoff’s face-displays and Kiviat graphs can effectively display more variables, they remain limited to bounded, numeric data and have never become widespread due to this and other limitations. Almost all intensive care units continue to use standard ‘flow sheets’ that are sometimes unique for each unit in a large hospital. Each unit thus develops their own specialized flowcharting methods for patient information, because a large majority of that data is the same from unit to unit (laboratory results, patient devices, X-rays, etc).

The design of the symbolic system (glyphs) we utilized is the knowledge-enhanced graphical symbols (KEGS). Individually, a KEGS is similar to a word in that it represents a concept, such as hemoglobin or hematocrit, which are then assembled into a sentence or set. For example, a complete blood count (CBC) is composed of data points or KEGS: hemoglobin, hematocrit, platelets, white blood cell counts and a differential white blood cell count. The set of KEGS are then arranged onto a human body background metaphor and displayed using a notepad device.

KEGS are rendered in colors and symbol shapes to represent varying degrees of high and low values, as well as normal. For example, a symbol that lights as a green rectangle with a red circle in the center represents slightly above expectations. If that symbol represented low-density lipoprotein (LDL) cholesterol then this would connote a mildly worse than expected condition. If the LDL was at an expected level (e.g. 130 mg/dl for a particular patient profile), then the symbol would be lit solid green. If, on the other hand, the symbol represented high-density lipoprotein (HDL) and the indicator was slightly above expectations (green rectangle with red center circle), then this would connote a mildly better than expected condition for a particular patient profile (Fig. 2).

In addition, a number of application specific symbols are available to describe non-numeric medical information such as patient devices and cultures. The background display is a simple human body metaphor upon which various data points are ‘plotted’ in relatively anatomical locations grouped into symbol sets, which are pictures comprised of multiple symbol sets (Fig. 3).

To manipulate the symbolic representations, instead of specific valuations being assigned to a given data point, such as Sodium of 142, a rule-based system translates all ‘normal’
Sodium values into the ‘normal’ symbol. Thus, if all the electrolytes fell in the normal range, the entire electrolyte symbol set would be colored green. The symbols allow the assignment of mildly high, moderately high, severely high and panic high (as well as low) values for any numeric data point. The ‘high’ symbols contain increasing amounts of red, while the ‘low’ symbols contain increasing amounts of blue and panic high and low are represented by flashing red and blue rectangles containing ‘!’, respectively. By using plot location, relative to the anatomical metaphor background as a data label, even signs or symptoms, for example headache or diarrhea, can be shown on the same display, using the same paradigm.

Figure 2 Knowledge enhanced graphical symbols.

Figure 3  GIFIC KE GSets display.
Methods

Objectives

Since studies [24] have shown that visual representations can be structured to reduce linearity and enhance comprehension of information, we proposed that cognitive load should be reduced with a symbolic representation with transformational grammar characteristics (knowledge enhanced graphical symbols), and that this reduction in cognitive load would have a positive impact on physician performance. The KEGS focusses on information extraction by facilitating perception and recognition of patterns for rendering patient information compared to more traditional forms of information representations, which rely heavily on cognitively intensive memory, integration and inference capabilities [25]. To investigate this, we conducted two exploratory studies; the first using a dual-task test comparing cognitive loads with KEGS compared with equivalent traditional forms of information representation used in a hospital intensive care unit. Following from this study, we then assessed physician performance on time and accuracy of diagnostics.

Cognitive load experiment

The dual-task methodology is based on the assumption that the processing capacity of working memory is limited, but can be flexibly allocated [31]. If two tasks have to be processed at the same time (dual-task condition), and both require the same cognitive resources, then these resources have to be split between the two tasks. This means that fewer resources are available for processing each individual task than would be available for processing a single task (single task condition). If the processing of a task depends on available cognitive resources, then performance in processing a secondary task will be reduced in relation to the amount of cognitive resources required by the primary task [18]. Here it is assumed that different variants of primary tasks require different amounts of cognitive resources. Thus, the performance in processing a simultaneously presented secondary task varies according to the cognitive load induced by the primary task. The dependent variables in these tasks are reaction time (interval between task presentation and reaction) and error rate pertaining to the processing of the secondary task. In the measurement of cognitive load in the processing of visual forms of information, we are interested in dissecting automaticity from intentional uses of cognition to determine cognitive load.

Participants and design

Participants consisted of 42 third- and fourth-year medical students (24 male, 18 female) at three medical schools in the United States. Participants’ ages ranged from 25 to 33 years. All of the participants had completed at least part 1 of the US Medical Licensing Examination administered by the National Board of Medical Examiners and had begun working in intensive care specialties of medicine. Since the dual-task test measures cognitive load, accuracy in diagnoses that depend on expertise was not a concern. Since our focus was on measuring cognitive load and not accuracy, student populations were appropriate for this test.

Nevertheless, task performance has many individual variables, such as prior knowledge and cognitive capabilities. However, when performance on a task is measured for a given individual, these effects can be controlled in the research design. Owing to expected large individual differences in reaction times, an all-within-subject design with repeated measures was used, and pretests were controlled as a covariate. Dependent variables consisted of reaction times for the secondary task and knowledge acquisition for the primary task (interpreting the information). A pretest was performed to allow for the assessment of prior knowledge effects for the primary task. To achieve a baseline performance level, participants were given a battery of tests that included both traditional charts and KEGS representations. Gain scores were used to determine learning trails. When the gain scores were no longer statistically significant, a baseline was established for each individual for the pretest–posttest conditions. Because the conditions ranged in type and severity, the response times were averaged, and their baseline performance measures were used as covariates in the analyses.

Methodology

Brunken et al. [9] constructed a dual-task test to measure learning under cognitive load with multimedia. This study provided a model for our study of cognitive load to interpret information in various forms. Using their model, for the secondary task, a simple, continuous, visual observation task was used in which participants were presented with a series of 50 different patient conditions in traditional form and in KEGS form. The secondary task used a colored letter ‘4’ displayed in a separate section of the window on the screen of the display. After a random period of 5–10 s, the color of the letter was changed (e.g. from black to red). The participants were to press a designated key on the keypad as soon as possible after the letter had changed the color. Once the key was pressed, the response time was recorded and the next countdown started. The software automatically recorded the lapse between the appearance of the letter in a new color and the key press. For the analysis of reaction times, the data from the secondary task were first synchronized with the program on the basis of time-stamped log file data. Then the secondary task measures were matched to their corresponding test condition: chart interpretation primary task versus KEGS interpretation primary task.

Results

The data were analyzed using one-way ANCOVA with pretest-baseline scores as the covariate to control for learning trials and individual performance differences. The assumption of homogeneity of regression slopes was tested in the interaction between test conditions and pretest baseline.
scores. The interaction was not significant \( (P = 0.79) \), indicating the assumption of homogeneity of regression slopes was tenable. The test of conditions versus response times to the secondary task in clock seconds time showed significant reduction \( (P = 0.001) \) in cognitive load for the KEGS (clock seconds mean = 0.14, SD = 0.15) compared with the traditional information (clock seconds mean = 0.53, SD = 0.19) (Table 1).

**Performance experiment**

In Experiment 2, we were interested in digging deeper into the possible effects of cognitive load on time and accuracy in interpreting patient conditions, and therefore expertise became a factor. In this experiment, we solicited board-certified intensivists to compare their performance with standard charts and flow sheets versus the KEGS.

**Participants and design**

Three board-certified expert intensivists trained in the KEGS participated in the exploration of evaluation of time and accuracy; two were compared their performance, while the third acted as a judge. There were two stages for this test. In the first stage, participants were told that time was more important than accuracy. For this test, both participants evaluated 10 cases each (5 in standard form and 5 in KEGS form). For the second stage, participants were told that accuracy was more important than time. For this test, both participants evaluated 30 cases each (15 in standard form and 15 in KEGS form).

**Methodology**

Participant physicians were evaluated by the judge on their assessments of patient information, and then had to answer a standardized questionnaire regarding diagnosis, physiologic abnormalities and level of illness and treatment plans regarding the patients. The questionnaires were then independently reviewed by the judge for accuracy, who scored the results without knowledge of which questionnaire was filled out by which participant. Accuracy scores were on a scale of 1 to 10, with 10 being perfect accuracy. In both instances, all patients were picked at random by the nurse manager in the critical care unit and were unknown to the reviewing physicians. No patient examination was allowed; however, the traditional chart and flow sheet physician had access to all patient progress notes, histories and physicals, etc. The KEGS physician had access to a single display.

**Results**

When time and accuracy were compared under the condition that more emphasis was placed on time than accuracy, time (in clock seconds) was better \( (P < 0.000) \) for the KEGS interpretation (clock seconds mean = 4.32, SD = 0.48) than that for the traditional rendering (clock seconds mean = 9.73, SD = 3.21), but was not significant \( (P = 0.136) \) for accuracy on the judge’s scale of 10 (KEGS, accuracy mean = 8.70, SD = 0.49, traditional, accuracy mean = 8.00, SD = 1.33). Hence, there was no difference in accuracy when emphasis was placed on time.

For the second test where more emphasis was placed on accuracy rather than time, at the 0.05 level both time (KEGS, clock seconds mean = 3.59, SD = 1.57, traditional, clock seconds mean = 9.77, SD = 4.16, \( P < 0.000 \) ) and accuracy on the judge’s scale of 10 (KEGS, accuracy mean = 8.81, SD = 0.75; traditional, accuracy mean = 8.28, SD = 0.94, \( P = 0.022 \) ) were statistically better for KEGS than for the traditional rendering.

**Discussion and limitations**

Because technologies may ease the demands placed on healthcare providers, we studied the effects on cognitive load and performance in working with KEGS compared to traditional forms of patient information. The KEGS appear less cognitively taxing than working with conventional charts and graphs in the diagnosis of patient conditions. When applied in a practical setting, these symbols may require less time for physicians to make accurate diagnoses. The theory of cognitive load also suggests that there should be a proportionate increase in errors as the cognitive demands elevate, for example in times of crises or patient influx. The reduction in the cognitive load levied by information rendering should mitigate, and should be considered by healthcare technology designers. Our study indicates that more emphasis should be placed on how information is presented to physicians by reducing cognitive load. Although enhancements could be made to the traditional chart and flow sheets, moving toward more simultaneous displays of information should enhance physician capabilities.

<table>
<thead>
<tr>
<th>Table 1 Means and Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Condition</strong></td>
</tr>
<tr>
<td>Charts ( (n = 42) )</td>
</tr>
<tr>
<td>KEGRS ( (n = 42) )</td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
</tr>
<tr>
<td>Obtained Mean</td>
</tr>
<tr>
<td>0.54</td>
</tr>
<tr>
<td>0.14</td>
</tr>
</tbody>
</table>
Given the practical problem enlisting practicing physicians in research studies, for the cognitive load experiment we utilized student populations; hence we were unable to use accuracy in diagnoses as a control variable, because it was contaminated by domain expertise. However, since our focus was on cognitive load and we were able to control for individual differences and individual learning, we were able to observe a reduction in cognitive load. Nevertheless, we qualify our findings as exploratory. To move beyond the exploratory phase, future research should utilize randomized experiments with larger populations, as well as increase ecological validity by applying the types of demands typically found in emergency or intensive care units. Also, future research may consider similar study using the mental effort rating scale [32].

References