Individual differences and task-based user interface evaluation: a case study of pending tasks in email

Jacek Gwizdka*, Mark Chignell

Interactive Media Laboratory, Knowledge Media Design Institute, Mechanical and Industrial Engineering, University of Toronto, 40 St. George St, BA 8171B, Toronto, Ont., Canada M5S 2E4

Abstract

This paper addresses issues raised by the ever-expanding role of email as a multi-faceted application that combines communication, collaboration, and task management. Individual differences analysis was used to contrast two email user interfaces in terms of their demands on users. The results of this analysis were then interpreted in terms of their implications for designing more inclusive interfaces that meet the needs of users with widely ranging abilities.

The specific target of this research is the development of a new type of email message representation that makes pending tasks more visible. We describe a study that compared a new way of representing tasks in an email inbox, with a more standard representation (the Microsoft Outlook inbox). The study consisted of an experiment that examined how people with different levels of three specific cognitive capabilities (flexibility of closure, visual memory, and working memory) perform when using these representations. We then identified combinations of representation and task that are disadvantageous for people with low levels of the measured capabilities.
© 2004 Elsevier B.V. All rights reserved.

1. Introduction and motivation

The construct of usability is generally considered with respect to a broad cross-section of people. This broad notion of usability may subsume modified levels of usability for specific individuals when using specific artifacts (e.g. software interfaces) to carry out

* Corresponding author. Tel.: +1-416-946-5836.
  E-mail addresses: elsevier_author@gwizdka.com (J. Gwizdka); chignell@mie.utoronto.ca (M. Chignell).

specific tasks. For instance, a virtual reality application might be considered to be reasonably usable for part of the user population, but not for those with low levels of spatial ability (Modjeska and Chignell, 2003). Dillon and Watson (1996) suggested that one goal of user interface design may be to reduce differences between users. This may be done by identifying the user characteristics that predict the biggest differences in performance. Sources of variation at the task or interface component level may then be isolated, and task or interface redesigned to minimize the offending components (Egan, 1988). The goal of such analysis and redesign is to maximize benefits of using the interface/system for all user groups.

In this paper, we use this type of approach with respect to a comparison of two alternative user interfaces to Email inboxes. Three user characteristics (visual memory, working memory, and flexibility of closure) were measured along with the effects of these characteristics on task performance in the two interface conditions (labelled as 'UI-Text' and 'UI-Visual'). The two tasks considered on each of these interfaces were a 'Date' task and a 'Header' task, as described later in this paper.

Our goal was to develop and test a methodology for evaluating email interfaces. In particular, we examined an interface intended to enhance management of pending tasks embedded within email messages. A secondary goal was to facilitate the design of more inclusive user interfaces by identifying and reducing differences in performance (attributable to differences in cognitive abilities) between users.

2. Related work

2.1. Multi-tasking and interruptions

For many years, personal computers have enabled users to perform multiple activities at the same time. Support was provided at the operating system level and encouraged at the user interface level by the introduction of windowing systems. The flow of interleaved user activities was studied in the context of human–computer interaction (Cypher, 1986). Psychological explanations of the phenomena related to managing multiple tasks and interruption handling were sought (Miyata and Norman, 1986).

External sources of human interruption on computing tasks increased dramatically with the growing use of email and instant messaging. The growing complexity of personal computing environments led to a recent upsurge of interest in the problem of how to support multi-tasking and handling interruptions at the user interface level. McFarlane and Latorella (2002) described a taxonomy of human interruption, and gave a survey of methods used in computer applications for coordinating interruptions. They also gave general user interface design guidelines for the support of three phases of human interruption (before, during, and after the task switch phase). Czerwinski et al. (2000) described interruption effects of instant messaging on task performance. Disruptive effects of interrupting messaging on long term memory were demonstrated by Oulasvirta and Saariluoma (2004). Czerwinski et al. (2004) described what prompts computer users to switch tasks, and the difficulties they have in resuming interrupted tasks. The findings were used to create design guidelines for task-management tools.

2.2. Email

Email is not just an asynchronous communication tool. It has become an environment for conducting work, and for maintaining social life. It is used for file transmission, task management, team collaboration for managing professional and social contacts, and, for conversations. This diversity of email use has been observed in numerous studies (Mackay, 1988; Whittaker and Sidner, 1996; Ducheneaut and Bellotti, 2001; Gwizdka, 2001) and is now widely recognized. Yet, this variety of activities performed in email is not reflected in email interfaces. Those have changed very little since email's invention.

In an early study, Mackay (1988) described how email supports a variety of time and task management activities. Whittaker and Sidner (1996) observed how the inbox is used as a repository of information containing to-dos, to-reads and other messages that cannot be dealt with immediately upon reading. Ducheneaut and Bellotti (2001) discussed how email becomes the central place where work is received, delegated, and managed. Most email clients employ the messaging metaphor that has not been designed for task management, and that does not support users in performing those activities in email.

In principle, two approaches to supporting task management in email are possible. In the first, the support is provided directly in email. In the second, email is integrated with other information management tools. Several email interface prototypes have recently been developed to support task management in email. Bälter and Sidner (2000) employed a pile metaphor to support tracking of tasks in email. Bellotti et al. (2003) created TaskMaster, providing probably the most comprehensive support for task management in email to date. Based on the premise that the task, and not the message, is the main element of interest, TaskMaster introduces *thrasks*, which are threaded collections of messages. The main distinction from other work is that thrasks correspond not only to message threads, but also to collections of messages related to meaningful user activities. Thrasks are created semi-automatically. The TaskMaster interface provides ways to track and prioritize different tasks. An example of the second approach is presented by Kaptelinin (2003). In his UMEA system, email functionality is associated with a more general workspace, along with other functionality (browsing the web or local file system, document editing, printing). Instead of adding support for other activities to email, Kaptelinin's approach organizes different types of information around higher-level goals of the user. Kaptelinin's approach reflects a move away from application- and document-centric to goal-centric computing. This approach is informed by Activity Theory (Kaptelinin, 1996; Kuutti, 1996).

In addition to addressing task management in email, other new email interfaces are proposed to better show relationships among messages (Rohall et al., 2001; Sudarsky et al. 2002; Venolia and Neustaedter, 2003), to provide message summarization (Rohall and Gruen, 2002) and to exploit social relations among one's correspondents (Farnham 2002, Whittaker et al., 2002). Little is known, however, about how to evaluate these new interfaces and visualizations and how to measure the improvement attributable to them (Gwizdka and Whittaker, 2003). More research is needed that examines what role different email interfaces and tasks might have on performance and other observed effects.

Due, perhaps, to a lack of established measures and methodologies, the evaluation of email interfaces with respect to their impact on task awareness is often limited to reports

based on subjective experiences. More research is needed that examines what role different email interfaces and tasks might have on observed effects. As an example of this research, Neuwirth et al. (1998) describe a task-driven design for email interfaces using the following steps:

(1) Develop benchmarks tasks (designed to capture aspects of real-world tasks);
(2) Observe subjects performing the benchmark tasks;
(3) Develop prototypes to overcome these difficulties;
(4) Measure and compare the performance of subjects completing the tasks with and without the aid of the prototype;
(5) Make changes to the prototype based on the results of the comparison, and repeat the process.

2.3. Individual differences in HCI—cognitive abilities

Effects of individual differences on human–machine interaction, and especially of cognitive ability, have been studied by the human–computer interaction community since the 1980s. Cognitive ability has been recognized as an important predictor of computer-based performance (for reviews see Egan, 1988; Westerman, 1993; Chen and Rada, 1996; Dillon and Weston, 1996). The reported differences in performance for computing tasks have been found to be quite large. For example, Egan (1988) reported differences between users in the order of 20:1 for performance of common computing tasks. Egan suggested that these differences could be predicted, and modified through appropriate design.

Subsequent research has considered the effect of a variety of individual cognitive differences on performance, including: spatial ability, visualization ability, spatial relations, closure speed, closure flexibility, perceptual speed, locus of control, WM, associative memory, visual memory (VM), associative memory, associational fluency, ideational fluency, and learning style. In particular, spatial ability has received considerable attention. Its effects on performance have been studied by many researchers, in areas and applications such as: navigation in virtual environments and in hypertext; textual information retrieval; visual information retrieval. For example, Westerman (1995) studied the effects of spatial ability on network navigation tasks. High spatial ability has been found to benefit users particularly in low-semantic contexts, when little semantic structure is applied to the network visualization. The effects of spatial ability and associative memory on performance in information retrieval in virtual environments were studied by Westerman and Cribbin (2000a,b). Users with high spatial ability had overall better performance. Similarly, Modjeska and Chignell (2003) found that people with low spatial ability had significantly slower performance when searching for information in a desktop virtual reality environment.

Other studies have demonstrated reduction of the performance gap between different population groups through appropriate design modifications. Sein et al. (1993) conducted a study examining effects of visual ability on the users' ability to learn three software applications. Use of a direct manipulation interface led to a reduced difference between high and low visual ability users in their study. Zhang et al. (Zhang and Salvendy, 2001) investigated the effects of users' visualization ability and website structure display design

on web browsing performance. They found that structure preview reduced the differences in performance between high and low visualization ability users.

According to human information processing models (Lindsay and Norman, 1977; Baddeley, 1986), WM plays a critical role as an input buffer for all information incoming from human senses. Limited capacity of WM is a well-known bottleneck in human information processing (Miller, 1956). Based on a review of a large set of research results, Miller posited a range of individual differences in the capacity of WM to be equal to seven plus or minus two, where the capacity limitation is with respect to the number of individual items, or to the number of meaningful groups (chunks) of items. The role of individual differences in capacity of WM in graphical information processing was shown, for example, by Lohse (1997).

3. Research strategy

Our research perspective assumed that task performance is aided by external and distributed cognition (Hutchins, 1995), where the external environment is seen as playing an integral role in cognition. It has been argued (Gruen, 1996; Zhang and Norman, 1994) that external representations of information can have a critical impact on how that information will be used in task execution. The role of external representations and their interaction with internal representations and processes is important for understanding how people manage multiple pending tasks. Malone (1983) argued that office desks are organized to remind people about these tasks. Spatial arrangement is used to represent activities, their priorities, temporal dependencies and relationships among multiple tasks. Management of everyday activities relies heavily on such spatial placement and manipulation of physical objects (Gruen, 1996). The same user habits can be brought to bear on the management of pending tasks in email, and, more generally, in the computer's desktop interface. The research questions that we would like to answer are:

- What external representations of pending tasks provide the best support for their management?
- How do individual differences affect the usefulness of different external representations?

4. Time in Email

Time is a necessary factor in the management of pending tasks. We outline below the design space of alternative email user interfaces supporting task management by describing the relationships between messages and time. Email messages may contain multiple references to time. The two most common temporal references contained in messages are *message arrival time* and *message reference time*. The message arrival time is always in the past, while messages can refer to the past, the present, or the future. These two timelines are embedded, explicitly or implicitly, in email messages.

4.1. Design strategies

The linear, tabular format of email folders has been designed to support a *one-touch model* of dealing with email messages (Whittaker and Sidner, 1996). In this model, once a new message has been read, the email user is expected to execute an immediate action which can be a combination of one or more of the following: (1) respond to the message, (2) delete it, or (3) file it. The message may later be retrieved from the archive, but it is not active any more. This model of processing, however, does not work for messages that cannot be acted upon immediately after their arrival. The problem is clearly apparent in the case of messages containing some type of future reference (e.g. messages carrying pending tasks). These messages need to be kept around, and, typically, they remain in the inbox. Email users thus perform the fourth possible action: (4) leave message in the inbox. The one-touch model breaks down when more and more messages are left in the inbox. The model supports past and present only, it does not support the future. Users are forced to repeatedly review messages left in their inboxes. They cope with this constraint by employing a variety of strategies, such as limiting the reviewing process to one screen full of messages.

Such issues can be ameliorated by using the two timelines embedded in email messages to support their management. The message arrival timeline can be used to facilitate retrieval of messages by bringing human autobiographical memory (Conway, 1990) to bear on retrieval of messages. The message temporal reference timeline can be used to support the management of messages referring to the future (containing pending tasks). In the case of each timeline, user interface can be designed to make temporal relationships between email messages and their temporal attributes perceptually explicit, thus shifting some of the load from the user's cognitive system to the perceptual system (Robertson et al., 1991). The underlying premise is that appropriate visual representations maybe more usable than textual representations. The timelines can be represented using a common representation of a horizontal line with the time arrow running from left to right. The vertical axis in such a visual representation can be used to distinguish between email messages by means of another attribute, for example, one of the header fields from email message.

TimeStore, a novel, time-based email interface proposed and developed by Yiu et al. (1997), is an example of the first approach using message arrival timeline. Messages are automatically organized by time and by sender, and are displayed on a two-dimensional grid (Fig. 1).

The two-dimensional representation allows messages to be located by using cues from autobiographical memory: when the message was received, and by whom it was sent. Autobiographical memory is a memory for events, which we have experienced (Conway, 1990). TimeStore users liked the visualization of their email and found it useful for retrieval of both old (*inactive*) and new (*active*) messages (Jovicic, 2000). The use of message received time in TimeStore to organize email messages is similar to Outlook's feature called *Journal*. However, in Outlook's Journal email messages are displayed on a horizontal linear timeline and there is no further organization of messages provided.

The second type of timeline, the *temporal reference of messages*, is used to facilitate the management of pending tasks embedded in messages in TaskView, the interface used in
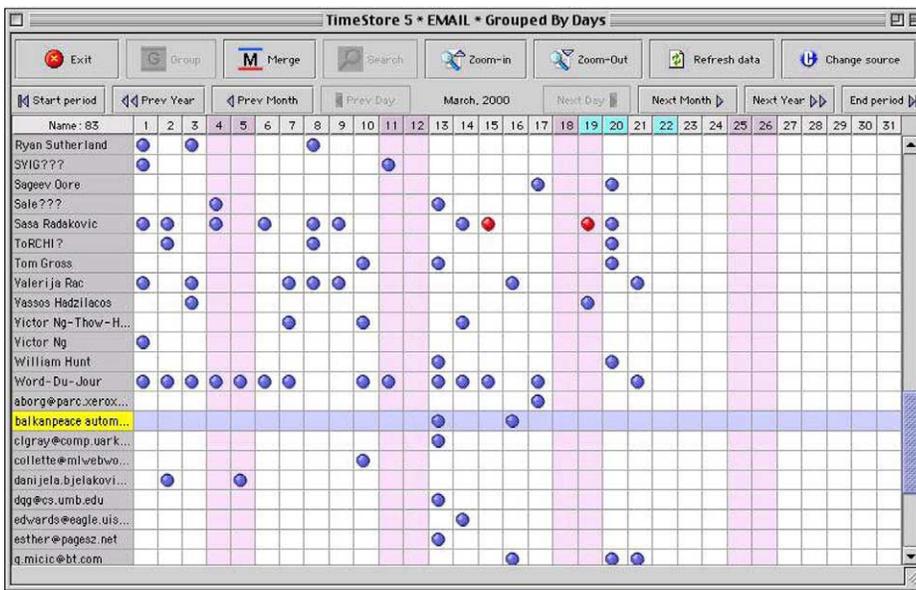
Fig. 1. TimeStore interface (Yiu et al. 1997)—one month view.

the study described in this paper. The TaskView is based on TimeStore and uses the same graphical representation (Fig. 2). In TaskView, tasks embedded in messages are represented by small icons on a two-dimensional grid with temporal and other task information shown on the horizontal and vertical axis, respectively. Other task attributes include sender, subject, or keywords extracted from the message body (user selectable). Navigation back and forward in time is provided. The displayed time period can be between 1 day and 1 year. The message body can be viewed by double clicking on the corresponding task icon.
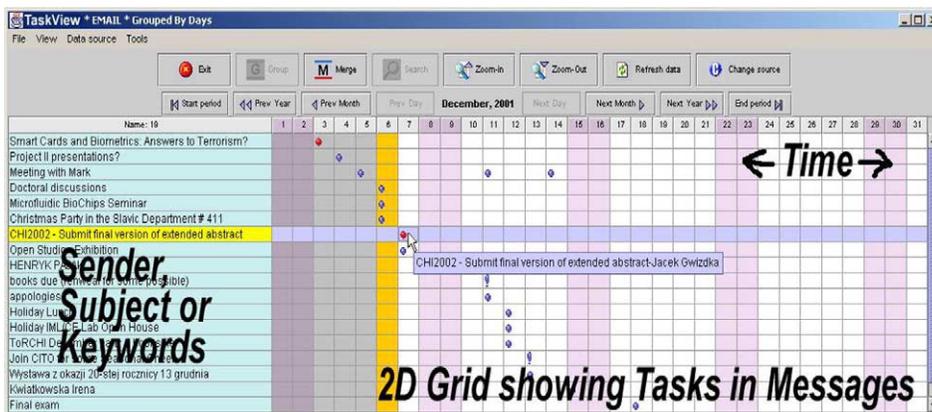

Fig. 2. TaskView interface (UI-Visual experimental condition). Shown is monthly view with pending tasks sorted by time.

Table 1
Main differences between external representations containing past or future references

Characteristic feature	Past reference (e.g. TimeStore)	Future reference (e.g. TaskView)
Temporal reference	Past	Future
Typical timeframe	Years → today	Today → months
Typical number of items	10–1,000,000s	10–100s
Temporal-perspective	Own—autobiographical	Own and others'—not autobiographical
Retrieval goal	Archive and current	Pending tasks, deadlines, delayed intentions

The TaskView interface was designed based on earlier results obtained with TimeStore, and through informal iterative design with rapid prototyping and user testing of preliminary versions of the TaskView interface. The development of the TaskView interface was also constrained by the requirement that the timeline of pending tasks needed to be made more visible in the user interface.

The TaskView interface can represent multiple pending tasks associated with one message (an example of such a case can be seen in Fig. 2). However, currently, all such tasks would be displayed on one horizontal line and thus be associated with the same message subject line (and thus with one message body). Therefore, an initial simplifying assumption was made that one message corresponds to one task.

The main focus of this research is on the presentation of pending task information. Hence, we do not describe the process of extracting temporal task attributes from email messages. For the purpose of our study, date information was extracted manually. An automated approach to date extraction from email messages is described, for example, by Stern (2003).

TaskView presents only active messages, that is, messages with future references containing pending tasks. The presentation is in future time, referencing pending tasks, while in TimeStore the presentation is in past time, arranged according to message arrival time. Key differences between external representations containing past or future references are summarized in Table 1.

5. User study

A user study was conducted to examine the proposed (more visual) representation of pending tasks as implemented in the TaskView interface. A more textual email interface (Microsoft Outlook) served as a baseline email interface (Fig. 3). Messages used in the study contained pending tasks. In the Outlook condition those tasks contained messages were presented as textual fields (Follow-Up-By and Due-By) in a tabular view of the Inbox folder (Fig. 3). The 'Follow-Up-By' and 'Due-By' fields are not shown in the default Outlook Inbox folder view. These fields can be optionally added to any folder view in Outlook. Their contents need to be filled out manually. In the TaskView interface, pending tasks contained in messages were represented on a two-dimensional grid (Fig. 2).
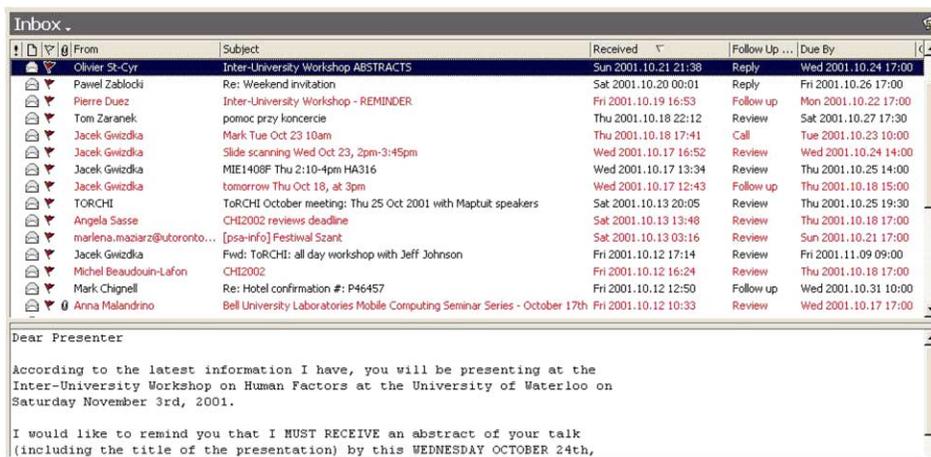
Fig. 3. Outlook Inbox interface—UI-Text experimental condition.

5.1. User task

To evaluate the TaskView interface, we selected information finding in email messages as a user task. Numerous studies of email use have demonstrated that email users frequently keep information in their inboxes and email folders. Findings from our earlier field study (Gwizdka, 2004) demonstrated that users who keep email messages containing future references (e.g. pending tasks), tend to visit their inboxes more often. These operations require email users to find information in messages. The selected user task is performed quite frequently in email handling sessions, often as a part of more complex email activities.

5.1.1. The metrics

The growing awareness of, and need for, effective and efficient email handling (Jackson et al., 2001) led us to use the following general metrics: (1) ability to find information in inbox (success of task completion); (2) accuracy of the answer; and (3) time to find the information.

As noted earlier, the main user task was to find information related to pending tasks in email messages. The task was designed to simulate a real email session in which users are looking for information in email messages based on partial information. For example, one may want to find out how many meetings (described in email messages) are scheduled for next week, or one may remember that there is a lunch in two weeks, but not the exact date of that lunch. The information finding task was driven by multiple-choice questions displayed on a computer screen, along with the choice of possible answers.

The questions were designed to refer to two different types of target information: (1) dates associated with pending tasks; and (2) sender or subject information contained in message headers. The two types of target information were chosen to compare performance on the two kinds of user interfaces. By design, presentation of the first type of information differed between the two user interfaces, while presentation of

the second type of information did not. Thus, we expected that answering questions about dates would involve different information finding strategy in the more textual interface (Outlook) than it would in the more visual interface (TaskView). In contrast, search for information in the message header was expected to be less affected by differences between the text and visual interfaces. The two information types were used to create two categories of questions and two corresponding information finding tasks ('Header', 'Date'). Type 'Header' questions refer to non-temporal information in the message subject, or sender, header fields. Type 'Date' questions referred to pending tasks' temporal information, for example, to a meeting date, to a to-do deadline, or to the number of events scheduled in a specified period of time.

A set of 49 questions about pending tasks contained in the messages was created from a set of queries that people ask their calendars and to-do lists. This set of queries was generated based on information collected from two people (one was a manager of an architectural department at a major national bank; the second was a manager in a large telecommunications company). As described earlier, the questions were categorized into two types: 'Header' and 'Date'. The questions are listed in Appendix B.

5.2. Cognitive abilities used in the study

Based on the user interfaces used in our study and on the nature of tasks in our study we choose to study three cognitive abilities: *flexibility of closure*, *visual memory* and *working memory*. These factors were identified as distinct cognitive abilities. They belong to a larger set of commonly agreed upon primary cognitive factors (Cattel, 1974; Ekstrom et al., 1976; Carroll, 1993), which are listed in Appendix A. Below, we give definitions of the three selected factors after Ekstrom et al. (1976), along with our rationale for selecting them.

- Flexibility of closure (FC)
  Definition. (Ekstrom et al., 1976). The ability to hold a given visual percept or configuration in mind so as to disembed it from other well defined perceptual material.
  Rationale. We used FC because finding an email message requires first locating the message representation (graphical or textual) among other email messages, and then holding the message representation while checking other messages.

  Processing information in a graphical email user interface requires extracting email message or email message attributes from a distracting background of other messages, which requires an ability to extract parts from the whole. This ability is referred to as FC. Carroll (1974) described FC as 'a process occurring in WM whereby a figure is imagined in elation to a surrounding visual-representation field'.
- Visual memory (VM)
  Definition. The ability to remember the configuration, location, and orientation of figural material.
  Rationale. Configuration and location of visual representations of email message needs to be remembered by users when moving between screens, and when scrolling within windows.

  Prior research studies have demonstrated reduction of the performance gap between different population groups through appropriate design modifications. Sein

et al. (1993) conducted a study examining effects of visual ability on the users' ability to learn three software applications. Use of a direct manipulation interface reduced the difference between high and low visual ability users in their study. Zhang and Salvendy (2001) investigated the effects of users' visualization ability (Ekstrom et al., 1976) and website structure display design on web browsing performance. They found that structure preview reduced the differences in performance between high- and low-visualization ability users.

VM is related to iconic memory and to short-term retention ability for visual material. VM involves different cognitive processes from those employed in other memory factors (Ekstrom et al., 1976).

- Working memory (WM)

    Definition. The ability to recall a number of distinct elements for immediate reproduction.

    Rationale. Currently dominant cognitive models posit that WM mediates all information incoming from human senses. This construct/ability appears particularly salient in email processing where the user has to organize, select and integrate large amounts of text.

According to human information processing models (Lindsay and Norman, 1977; Baddeley, 1986) WM plays a critical role as an input buffer for all information received through the senses. The limited capacity of WM is a well-known bottleneck in human information processing (Miller, 1956). The role of individual differences in WM capacity in graphical information processing was shown, for example, by Lohse (1997).

5.2.1. Relationship between cognitive ability factors

The cognitive ability factors cited above involve (at least partially) separate underlying cognitive processes. However, they are not to be understood as fully independent constructs. For example, FC involves WM processes. The cognitive ability factors provide indirect measures of underlying cognitive abilities.

5.3. Hypotheses

The first two experimental expectations were based on the belief that visual representation of pending task information will be beneficial to email users, in particular on the 'Date' task. Two specific hypotheses were formulated:

UI hypothesis. Efficiency as measured by performance time to complete information retrieval tasks will be higher overall in the UI-Visual (TaskView—Fig. 2) than in the UI-Text condition (Outlook—Fig. 3).

Task hypothesis. Performance on the 'Date' task will be faster in the UI-Visual than in the UI-Text condition.

It was expected that high levels of cognitive ability would generally increase the efficiency of pending task information retrieval, but that the effect of cognitive ability would be reduced with the UI-Visual interface. Thus three specific hypotheses were formulated (one for each of the cognitive abilities):

Flexibility of closure hypothesis. Participants with low level of FC will perform worse (in terms of efficiency) on the 'Date' task in UI-Text (Outlook) than in UI-Visual (TaskView). People in low FC group were expected to be slower on the 'Date' task in the UI-Text interface because this task required them to do extract textual task information from among other textual information. There was no easy perceptual/visual distinction between relevant (task) and non-relevant (non-task) information, both of which were presented in text.

Visual memory hypothesis. Participants with low VM will perform worse (in terms of efficiency) in UI-Visual (TaskView) than in UI-Text (Outlook). People with low VM were expected to be adversely affected by the more visual interface, because it required them to switch between different views (e.g. to navigate in time), and thus to hold more visual information in memory than when using the UI-Text interface.

Working memory hypothesis. Participants with high WM will perform in UI-Text (Outlook) better than those low on WM, while no such differences will be observed in UI-Visual (TaskView). This hypothesis derives from the role that WM serves as an input buffer and from the expectation that the Outlook interface requires more information to be kept in the input buffer than does UI-Visual.

5.4. Method

A mixed factorial design was employed with user interface as an independent within subject factor (two levels: UI-Text (Outlook) and UI-Visual (TaskView)). There were two sessions. Each subject used a different interface in each session. The design was balanced with respect to the order of interface use. There were also three independent, between subject factors, based on three measured cognitive abilities (FC, WM, and VM). Scores for each of the cognitive abilities in the experimental sample were split at the median into two groups (i.e. low vs. high levels of the ability). For each of the three cognitive abilities, these two levels of the ability (low vs. high) were defined as a pseudo-factor in subsequent analyses of the experimental data. There were two types of user tasks 'Header' (H) and 'Date' (D) as described earlier.

5.5. Apparatus

Two email programs were used in the experiment: Outlook (UI-Text) and TaskView (UI-Visual). The user interfaces for these programs, Inbox and TaskView, respectively, were described in Section 4.1. The programs were installed on a desktop PC in the experimenter's office. Participants' interaction was recorded using the Camtasia software for capturing activity on the computer screen. Email inboxes in both programs were populated in both sessions with the same 44 messages. The messages contained pending tasks, and were selected from a larger corpus of emails. This corpus was created from real-life messages by removing identifying information that would indicate who was sending, or being referred to, in each message.

5.6. Participants

Twenty-one subjects participated in the experiment. Eighteen participants were university graduate students (seven Master students and 11 PhD students) and three participants were full-time employees from outside companies or government agencies. There were seven females and 14 males. Participants were screened for at least moderate use of email and for the use of email to receive task information. On average, participants had used email for 6 years. Participants were paid $30 for their time ($10 per hour).

5.7. Procedure

The study consisted of four on-line questionnaires and two sessions conducted in the experimenter's office. The sessions were spread at least 2 days apart (2–7 days). Participants used a different email interface in each session.

Before coming to the first session, participants filled out an on-line survey containing demographic and email-habit questions. Each session began with explanations of the study protocol, followed by user interface training, and then the main task. The main task was to find information about pending tasks in email messages. Information finding was driven by multiple-choice questions displayed on screen. Twenty-one questions were drawn randomly in each session (with none of the questions being repeated from one session to the other) from a larger set of 49 questions (listed in Appendix B). After the main task was completed, a couple of cognitive tests were administered. Different tests were administered in each session. The tests used are listed in Section 5.8. In the 'TaskView' session, after using the new interface, participants filled out a subjective preference questionnaire. At the end of each session, participants were asked to freely recall information about pending tasks, which they had looked up in email messages earlier in the session. After each of the sessions participants filled out an on-line questionnaire, containing the same set of questions that they answered during the session (order of questions was randomized).

5.8. Measures

Experimentally controlled measures:

1. Two levels of user interface: UI-Text (Microsoft Outlook) and UI-Visual (TaskView);
2. Two tasks: Header (H) and Date (D).

Original independent, between subject measures:

1. Cognitive abilities were measured using the Factor-Referenced Kit of Tests (Ekstrom et al., 1976). FC was assessed using the CF-2 test; two VM tests were administered: (1) VM for shapes was assessed using the MV-1 test; (2) VM for object location on two-dimensional maps was assessed using the MV-2 test. WM was measured using the auditory digit span test (MS-1). The range of scores for the study population is shown in Table 2.

Table 2
Range of cognitive abilities for the study sample

| | Cognitive ability test scores | | | |
	CF-2[a]	MV-1[a]	MV-2[a]	MS-1 (WM)
Mean	58.7%	69.5%	65.8%	6.4	
Std. dev.	13%	16%	30.5%	0.9	
Min. value	32%	44%	0%	4.5	
Max value	80%	100%	100%	8.0	
Total number of participants	Low[b]	10	10	10	9
	High[b]	11	11	11	12
Participants who first used UI-Text	Low[b]	6	4	5	6
	High[b]	5	7	6	5
Participants who first used UI-Visual	Low[b]	4	6	5	3
	High[b]	6	4	5	7

[a] Test results are reported as percentage scores.
[b] Number 'Low' and 'High' are the numbers of people in the study sample who were, respectively, below or above the median value of each cognitive ability. The numbers are also given for each group of participants.

2. Demographic data and self-reported email use data, with a focus on handling of pending tasks, was collected using an on-line survey.

Correlations were checked for each pair of cognitive abilities test scores used in the study. Only the scores on MV-1 and MV-2 tests were found to be significantly correlated ($r = 0.638$, $p < 0.05$).
Derived independent, between subject measures (treated as pseudo factors):
From the Factor-Referenced Kit of Tests:

1. FC—scores on CF-2 grouped into two levels by median (low–high)
2. VM1—scores on MV-1 grouped into two levels by median (low–high)
3. VM2—scores on MV-2 grouped into two levels by median (low–high)
4. WM—scores on MS-1 grouped into two levels by median (low–high).

Dependent measures:

1. Efficiency, as measured by the time taken by participants to answer questions;
2. Effectiveness, as measured by ability to find the correct information (answer);
3. Subjective evaluation of both interfaces.

6. Results

The analysis began with an examination of the three-way ANOVA interactions between each of the three individual difference factors on the one hand, and the UI

Table 3
Order effect of experimental sessions (UI) on performance time in seconds

	First session	Second session
UI-Text	35	33.7
UI-Visual	37	30

and Task factors on the other. None of the three-way interactions was significant ($F \sim 1$ for short term memory, $F < 1$ for both FC and VM).

6.1. Order effect

There was a significant order effect of experimental sessions on performance time ($F(1, 19) = 6.6$, $p = 0.019$, $\eta^2 = 0.26$). Participants who used the visual interface first performed the tasks slower (37s per task on average), compared to the other three combinations of interface and its order (Table 3). The textual interface was familiar to all participants.[1] This effect may be thus attributed to having to learn the new, visual interface and the experimental task at the same time (in the first session). This hypothesis is supported by the data from the second session, where participants who used UI-Visual, that is, those who after learning the task in the first session were now learning the visual interface, performed slower than the other group in the second session (33.7 vs. 30s). In order to consider these effects explicitly, the following analyses report the results from the first and second session separately.

6.2. Effects of user interface design

The subsequent analyses were organized according to the study goals. First, we tested whether the design intervention incorporated in the UI-Visual (TaskView) interface was successful. (That is, whether study participants performed the experimental tasks more efficiently on the more visual interface.)

As explained above, the analyses were performed separately for each study session.[2] The two-way interaction between UI and Task was assessed using ANOVA. A significant interaction was found in each session (Session 1: $F(1, 17) = 5.2$, $p = 0.036$, $\eta^2 = 0.235$;[3] Session 2: $F(1, 17) = 11.8$, $p = 0.003$, $\eta^2 = 0.41$). The interaction appeared to be robust enough to have the same character in both sessions. The biggest difference was found for UI-Visual, where the 'Header' task required significantly more time (48.5s and 44.7s in first and second sessions, respectively) and was the slowest of all four UI*Task combinations, while the 'Date' task was the fastest (29s and 23.1s in first and second sessions, respectively) (Fig. 4). This interaction should be considered in the context of

---

[1] Sixteen out of 21 used Outlook (perhaps in addition to other email program), one used Eudora, and the remaining four used web-based mail. In all those cases, they were familiar with the textual email inbox.

[2] In the analyses of separate sessions that were carried out, UI became a between-subject factor.

[3] η^2 refers to the partial η (eta)—the proportion of variance that estimates the size of an effect (Murphy and Myors, 1998).
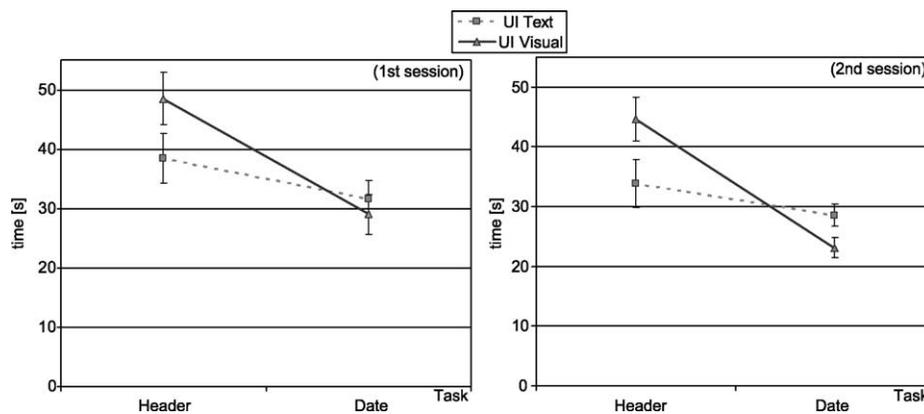
Fig. 4. Effects of interaction UI*Task on performance time in both sessions.

the main effect of Task, which was found as part of the same analysis. (Session 1: $F(1, 17) = 22.7$, $p < 0.001$, $\eta^2 = 0.571$; Session 2: $F(1, 17) = 32.35$, $p < 0.001$, $\eta^2 = 0.656$). The 'Header' task was significantly slower in both sessions (43.5s and 39.3s in first and second session, respectively), while the 'Date' task was faster in both sessions (30.3s and 25.9s in first and second session).

In the second session, when it could be assumed that the task-learning phase was completed, the 'Date' task was significantly faster in TaskView than in Outlook (23.1 vs. 28.6s). Therefore, the goal of 'better' supporting this type of task was achieved. In contrast, the 'Header' task was significantly slower in UI-Visual than in UI-Text (44.7 vs. 34s). This was an unforeseen result, as we expected that efficiency of performance in TaskView would not be worse than in Outlook. Possible reasons will be discussed later.

6.3. Effects of cognitive abilities

The effects of cognitive abilities were tested using separate two-way analyses of variance with the cognitive ability as one factor, and, either the type of interface or the task, as second factor. In the first case, it was a question of understanding the factors underlying the usability of an interface. Which population group will benefit from the interface? Which group will be affected adversely? In the second case, it was a question of understanding the demands of different tasks.

6.3.1. Working memory

The interaction between WM and task was found to be significant in the first session[4] ($F(1, 19) = 9.1$, $p = 0.008$, $\eta^2 = 0.349$), but not in the second. There was a significant difference in performance time for the 'Header' Task, with participants low on WM performing slower than participants high on WM (50 vs. 40s) (Fig. 5).

---

[4] There was also a similar effect for both sessions together. Both session effects (within subjects design) are not being reported here, because due to the order effect they are difficult to interpret (unless the same effect appears also in the first and in the second session separately).
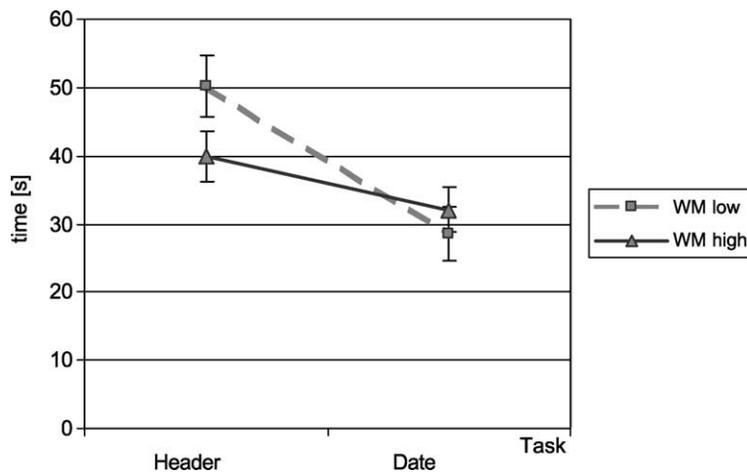
Fig. 5. Effect of interaction WM*Task on performance time in the first session.

A significant interaction was also found separately for the 'Header' task in the first session between WM and UI ($F(1, 17) = 4.8$, $p = 0.042$, $\eta^2 = 0.221$). Participants who were low on WM performed significantly slower (63s) in the UI-Visual interface (Fig. 6).

These two results taken together suggest that the level of WM affected performance on the 'Header' task, especially in the more visual TaskView interface. Since, these WM-related results appeared only in the first session, one can infer that they stem from the role of WM in learning the new TaskView interface and the 'Header' task.

6.3.2. Flexibility of closure

The interaction between FC and user interface was found to be significant in the second session (Session 2: $F(1, 17) = 6.32$, $p = 0.022$, $\eta^2 = 0.271$) (Fig. 7).
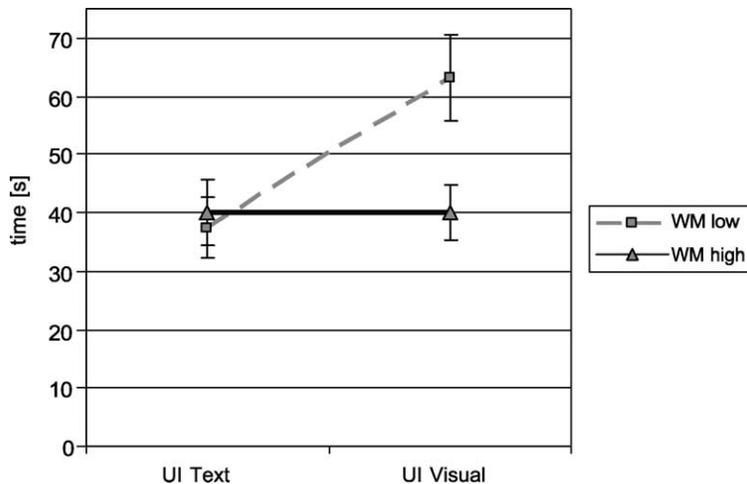

Fig. 6. Effect of interaction WM*UI on time in the first session (for the 'Header' task).

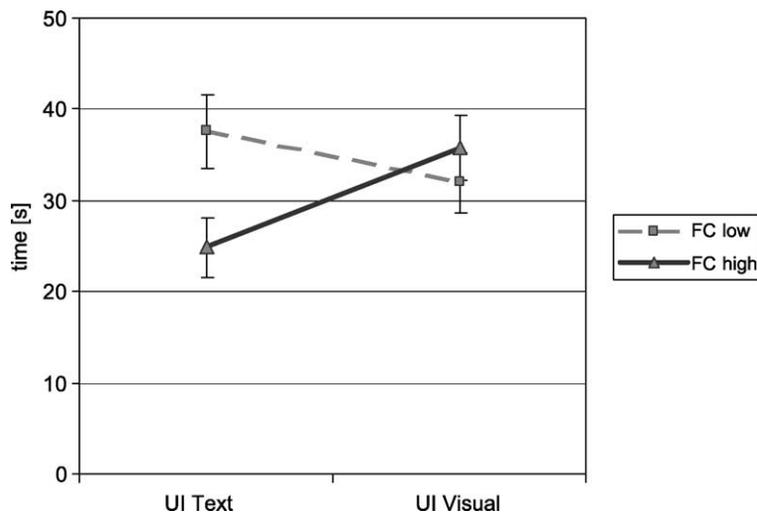*J. Gwizdka, M. Chignell / Interacting with Computers 16 (2004) 769–797*

Fig. 7. Effect of interaction FC*UI on performance time in the second session.

A significant interaction between FC and UI was found for the 'Date' task (Session 2: $F(1, 17) = 10.14$, $p = 0.005$, $\eta^2 = 0.374$) (Fig. 8). For the 'Date' task, there was also the main effect of FC on performance time (in both sessions), where participants low on FC were overall slower. In Session 1: low-FC 35.5s, high-FC 25.1s ($F(1, 17) = 5.15$, $p = 0.037$, $\eta^2 = 0.23$); Session 2: low-FC 28.8s, high-FC 22.9s ($F(1, 17) = 5.5$, $p = 0.031$, $\eta^2 = 0.244$).

As can be seen from Fig. 8, performance of people low on FC was adversely affected by the textual interface and the 'Date' task. Performance of people high on FC on the 'Date'
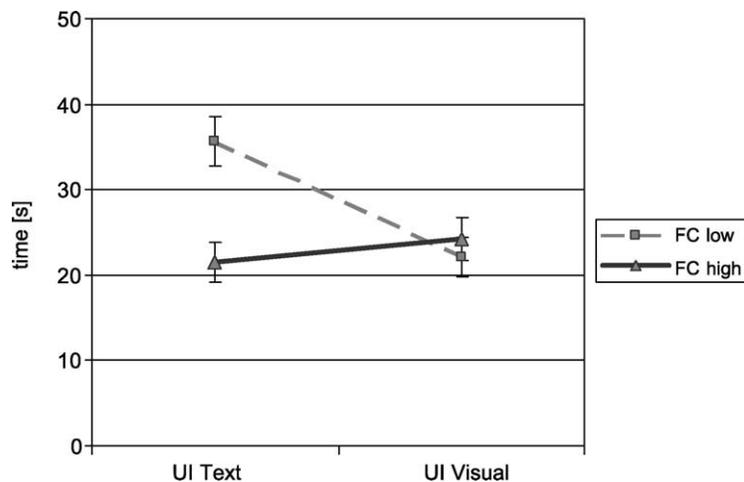

Fig. 8. Effect of interaction FC*UI on performance time in the second session (for the 'Date' task).
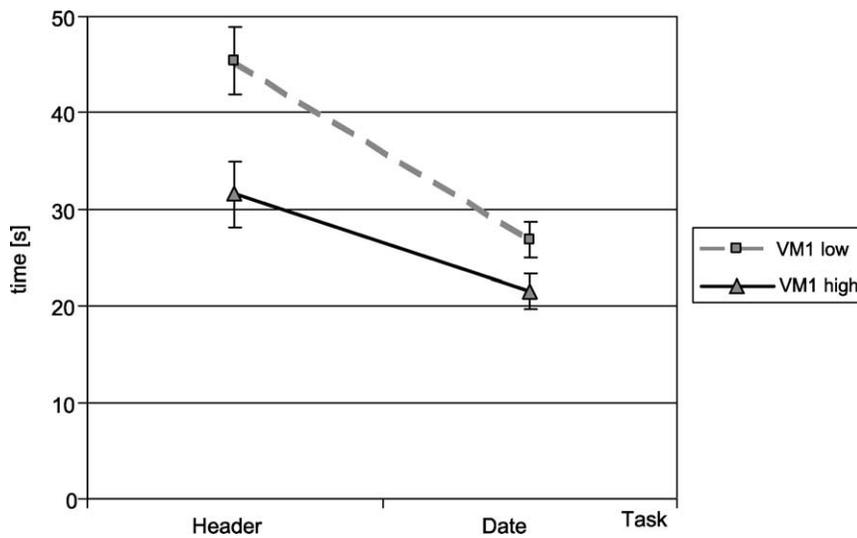
Fig. 9. Effect of interaction VM1*Task on time in the second session.

task was not affected. This result indicates that extracting date-related information is indeed easier in the more visual TaskView interface, which was designed for that purpose.

6.3.3. Visual memory

The interaction between VM measures (memory for shapes—VM1, and memory for location on a two-dimensional map—VM2) and task was found to be significant in the second session (for VM1: $F(1, 17) = 4.17, p = 0.058$ (borderline), $\eta^2 = 0.196$, see Fig. 9; for VM2: $F(1, 17) = 5.58, p = 0.030, \eta^2 = 0.247$ see Fig. 10).
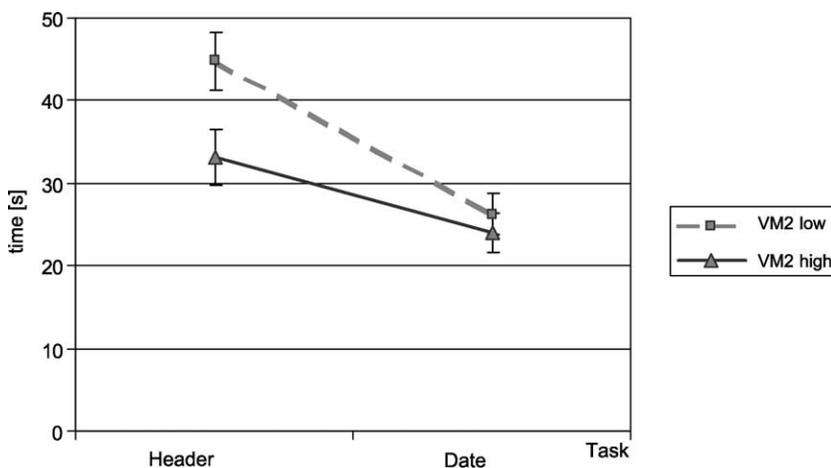

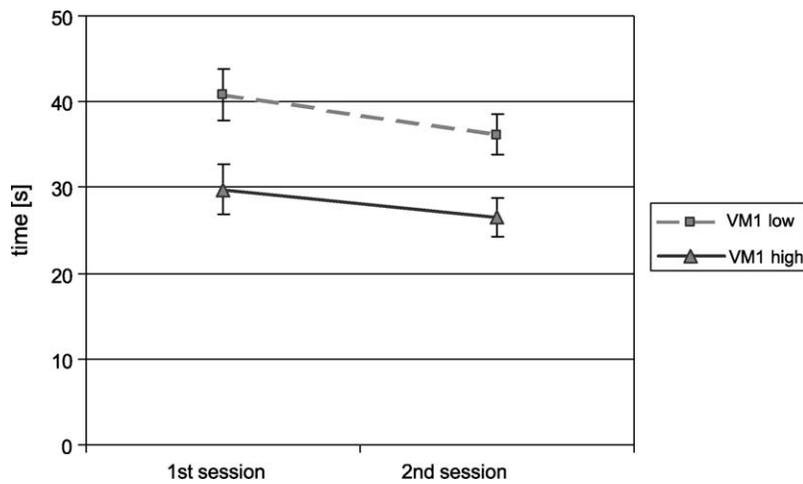Fig. 10. Effect of interaction VM2*Task on time in the second session.

Fig. 11. Main effect of VM1 on performance time in both sessions.

There was also a main effect of VM1 on performance time (Fig. 11), where participants low on VM1 were overall slower. In Session 1: low-VM1 40.8s, high-VM1 29.8s ($F(1, 17) = 6.5$, $p = 0.021$, $\eta^2 = 0.277$); Session 2: low-VM1 36.1s, high-VM1 26.5s ($F(1, 17) = 8.48$, $p = 0.010$, $\eta^2 = 0.333$).

As can be seen from Figs. 9 and 10, performance on the 'Header' task in the second session (after task learning was completed) was affected by the level of the two cognitive measures related to VM (VM1 and VM2), which were tested in the study. In both cases, people low on VM1 and VM2 performed slower than those high on VM1 and VM2. The 'Header' task required participants to switch between displaying sender vs. subject information.

6.4. Summary of cognitive ability effects

The main quantitative results are summarized in Table 4. In the first session, people with low WM were disadvantaged for the 'Header' task (H) in the TaskView interface. In the second session, people with low FC were slow on the 'Date' task (D) when using

Table 4
Summary of 'disadvantages' for people low on each tested cognitive ability

Cognitive ability	Disadvantage	
	First session	Second session
FC		Low FC—slow on Task D in UI-Text
VM		Low VM—slow on Task H
WM	Low WM—slow on Task H in UI-Visual	

the Outlook Inbox interface, while people with low VM were disadvantaged on the 'Header' task in both user interfaces.

In contrast to the results for performance time, there were no significant differences in accuracy across the various combinations of the experimental factors. The average accuracy across all conditions was 91%.

7. Discussion

7.1. Interface design and tasks

The new, more visual TaskView interface led to better performance for the 'Date' task (D). At the same time, this study showed the limitations of the current version of this interface. The benefits on performance time were observed for the 'Date' task only, while performance for the 'Header' task (H) was actually worse in TaskView than it was for the Outlook Inbox interface. The more general *UI Hypothesis* remained thus not confirmed, while the *Task Hypothesis* was confirmed. An observed order effect that was attributed to learning of the new TaskView interface and task, cannot fully account for this effect, since, on the 'Date' task in the first session, the visual interface was as good as the textual, and in the second session, the visual interface was better for the 'Date' task.

Thus one goal for the next iteration of the TaskView interface will be to redesign it to achieve at least the level of efficiency for the 'Header' task that is currently shown in a typical inbox interface. One simple modification that may contribute to achieving this goal is to display sender and subject information in two separate columns in the TaskView interface. This would avoid users having to switch between displaying sender vs. subject information. Such a design modification may be expected to allow performance at the level of the textual interface for 'Header' tasks and the TaskView interface for 'Date' tasks.

The TaskView interface can represent multiple pending tasks associated with one message. However, currently, all such tasks are displayed on one horizontal line and are associated with the same subject line (and with one message body). Thus, in the case of different (unrelated) tasks being described by one subject line, users may be forced to get more information by opening the message. Users with low WM and/or with low VM may be affected by the additional operations required. However, assuming one task per message should not reduce the validity of the study. Based on anecdotal observation, email users often wish to associate one task with one message. This tendency is reflected in the way they handle email messages containing multiple tasks. Their reply often addresses one task.

7.2. Cognitive abilities

Overall, high levels of each cognitive ability, as measured in the study, had a beneficial effect on performance time. The pattern of these effects suggests that different cognitive abilities affect different aspects of human–machine interaction. WM was found to affect learning (first session), VM to affect task performance (second session—after task

learning), while FC affected performance during and after learning (main effect of FC on 'Date' task performance in the first and second session).

The two user interfaces used in this study put different demands on different people. Users with low FC performed overall slower in the UI-Text (Outlook) condition. In particular, their performance on the 'Date' task was especially affected by the UI-Text condition, while it was not affected by the UI-Visual condition (confirming the *Flexibility Of Closure Hypothesis*, which expected participants low on FC to perform worse in Outlook than in TaskView on the 'Date' task). A possible explanation is that embedding messages (and the date information in particular) in the textual inbox, among other messages, may require more discrimination, which likely creates a disadvantage for people with low FC.

Participants with low VM performed worse for the 'Header' task than for the 'Date' task in both interfaces (not confirming the *Visual Memory Hypothesis*, which expected an interaction with the levels of user interface, such that participants low on VM would perform worse in TaskView than in Outlook). The 'Header' task required users to open and switch between more views (e.g. switching to display sender or subject information) and open more windows than in the 'Date' task. The switching caused changes in the visual field, which might have been difficult to cope with for people low on VM. Those people may perform better if more visual constancy is maintained (e.g. Woods, 1984).

The *Working Memory Hypothesis*, which expected participants with high WM to perform better than those low on WM in Outlook (but with no such differences expected to occur for the TaskView interface), was not confirmed. People with low WM had a greater disadvantage on the 'Header' task. The disadvantage was clearly visible in the TaskView interface, where performance of people with high WM was not affected at all. However, this significant difference appeared only in the first session. Thus, it may be attributable to the effect of heaving to learn both the new interface and a new task. The additional load imposed on WM by the need to hold more information when switching between different views (switching to display sender or subject information) did not significantly impact performance, since it did not appear in the second session. However, as discussed above, it did impact people low on VM.

7.3. User-centric reference-task-based methodology

Our user study results suggest that it is important to consider alternative interfaces for different population groups and for different tasks, i.e. to accommodate individual differences in ability. Egan proposed three steps in designing interfaces to accommodate individual differences (Egan, 1988), which are listed below:

1. *Ascertain* what user characteristics predict the biggest differences in performance.
2. *Isolate* the sources of variation at the task or interface component level.
3. *Redesign* tasks or interfaces to minimize their offending components so that the benefits for all user groups can be maximized.

Neuwirth et al. (1998) proposed a five step evaluative process for designing user interfaces so that important (benchmark) tasks could be carried out efficiently.

1. Develop benchmarks tasks (designed to capture aspects of real-world tasks);
2. Observe subjects performing the benchmark tasks;
3. Develop prototypes to overcome these difficulties;

Table 5
Comparison of approaches proposed by Neuwirth and Egan with the approach used in this paper

Approach	Neuwirth: task-centric	Egan: user-centric	Gwizdka and Chignell: user-centric reference-tasks-based
Goal	Develop or modify user interface to improve user performance on (a) specific task(s)	Develop or modify user interface or task(s) to minimize the differences in performance for all users	Develop or modify user interface to minimize the differences in performance for all users with respect to selected tasks
Steps	N1. Develop benchmarks task(s)		G1. Select and develop reference tasks and establish user characteristics (individual differences) that predict the biggest differences in performance
		E1. Ascertain what user characteristics predict the biggest differences in performance	
	N2. Observe users performing the benchmark tasks	E2. Isolate the sources of variation at the task or interface component level	G2. Study user performance, focusing on differences between user groups (e.g. characterized by different levels of cognitive abilities) at task and user interface level
	N3. Develop (or modify) user interface prototypes to overcome these difficulties	E3. Redesign task or interface to minimize their offending components so that the benefits for all user groups can be maximized	G3. Redesign user interface to minimize the differences between user groups across the selected tasks
		(Optionally) Repeat step E1.	(Optionally) Refine step G1 by adding or removing user characteristics that predict differences in performance
	N4. Measure and compare the performance of users completing the task(s) with and without the prototype user interface	Repeat step E2 (with modified task and/or interface).	Repeat step G2 (with modified task and/or interface).
	N5. Make changes to the prototype user interface based on the results of the comparison, and repeat the process	Repeat step E3.	Repeat step G3.

4. Measure and compare the performance of subjects completing the tasks with and without the aid of the prototype;
5. Make changes to the prototype based on the results of the comparison, and repeat the process.

Egan's framework is more general and, for example, does not prescribe how to isolate the sources of variation. Neuwirth et al. steps are more specific and suggest first observing the current situation (step 2) and then a comparative evaluation of before and after user interfaces (step 4).

In our research, we sought to combine designing for individual differences (Egan) with design strategy that enhances interface usability for key tasks (Neuwirth et al.). In Table 5, we summarize and compare these approaches and propose a combined approach, which aims to develop or modify the user interface to minimize the differences in performance for all users with respect to selected reference tasks (Whittaker et al., 2000).

In research carried out in this study, we followed the steps of our user-centric reference-tasks-based methodology (as presented in Table 5). Information finding in email inbox was selected as an important and realistic task. Based on the knowledge of the task, interface and human cognition, three cognitive abilities (individual differences) were selected that were likely to significantly influence task performance. A user study was conducted to assess the interaction between UI and Task and to isolate the sources of variation of user performance attributable to joint task/UI effects, and to task, and UI, effects separately. The results from this study can then inform the next design cycle, e.g. redesigning the interface to achieve better performance on the 'Header' task in the UI-Visual condition. We believe that this methodology is helpful in general, and can be used to accommodate individual differences in a variety of applications and user interfaces.

8. Summary

The prevalence and difficulty of email handling makes it an important target for usability enhancing re-design of user interface and functionality. Email messages containing future references are handled poorly in current email systems. This research examined how external representations of task information at the user interface can improve management and awareness of pending tasks that are encoded within email messages.

Specifically, this paper addressed the issue of enhancing the management of pending tasks in the email inbox. The research was designed to assess the impact of individual differences (specifically, three cognitive abilities) on the performance of selected tasks using each of two different email interfaces.

The interaction between UI and Task was assessed in an experiment to isolate the sources of variation of user performance attributable to a joint task/UI effect. This approach embodied a task-oriented approach to accommodating individual differences that combined the frameworks introduced earlier by Egan and by Neuwirth.

Users performed better on the 'Date' task in the TaskView interface than they did in the Outlook Inbox interface, while on the 'Header' task they performed better on the Outlook interface than they did in the TaskView interface. One design strategy for addressing this

trade-off in future will be to display sender and subject together in the left-hand column in TaskView. This would avoid users having to switch between displaying sender vs. subject information. We expect that such a design may allow performance at the level of the inbox interface for type 'Header' tasks and the TaskView interface for type 'Date' tasks. This mixed interface should also benefit people with low VM and FC by allowing them to choose to focus on the aspect of the interface that is most helpful or salient for a particular task.

The interaction between tasks and interfaces was independent of the three cognitive capabilities examined. However, there were significant interactions between the cognitive capabilities and the main effects of user interface and task. People with low VM were slower when using the TaskView interface and they were slower when carrying out the 'Header' task (relative to those with high VM). People with low WM were slower with the 'Header' task. These interactions identify combinations of user interface and task where users low on respective cognitive abilities were adversely affected.

These results suggest that individual difference effects pertaining to the use of email interfaces can be found with relatively little effort, providing user interface designers with a useful method for screening interface requirements and guiding general interface design strategies, as a supplement to existing usability engineering approaches.

The results of this study also highlight the importance of considering alternative interfaces, for different population groups and for different tasks, that would accommodate individual differences in ability. There appears to be considerable scope for improving email interfaces. Detailed analysis of the effects of different interfaces on users with different levels of cognitive ability will likely pay dividends in terms of informing new interface designs that can benefit a wide range of users.

Appendix A. The main primary factors of ability

Reprinted from Kline (2000) (Box 7.1).

V	verbal ability: understanding words and ideas
N	numerical factor: facility in the manipulation of numbers, not arithmetic reasoning
S	spatial ability: ability to visualize figures in different orientations
P	perceptual speed and accuracy: involving rapid assessment of differences between pairs of stimuli
Cs	speed of closure: the ability to complete a pattern with parts missing
I	inductive reasoning
Ma	rote memory: memory for pairs within which there are no mediating links
Mk	mechanical ability
Cf	flexibility of closure: ability to find stimuli embedded in distractors
Ms	memory span: the ability immediately to recall digits or letters
Sp	spelling
E	aesthetic judgment: the ability to detect the basic principles of good art
Mm	meaningful memory: the ability to learn links between pairs of linked stimuli

O1   originality of ideational flexibility: the ability to generate many different and original ideas
F1   ideational fluency, similar to O1 and O2: the ability to generate ideas on a topic rapidly
W   word fluency: rapid production of words conforming to letter requirements
O2   originality: marked by the test of combining two objects into a functional object
A   aiming: hand eye coordination
Rd   representational drawing ability
Au   auditory ability: the ability to differentiate between tones and to remember a sequence of tones.

Appendix B. Set of 49 questions used in the study

1. Do you have any tasks scheduled for the coming Saturday (Feb 22)?
2. At what time do you have tasks scheduled for the coming Saturday (Feb 23)?
3. Do you have more tasks scheduled for the next week (Feb 25–Mar 3) than for this week (Feb 18–24)?
4. Do you already have any tasks scheduled for the next month (March)?
5. Do you have any dinners scheduled for the next two weeks (-Mar 3)?
6. Do you have any days this week (Feb 18–24) without any tasks scheduled?
7. Do you have any days next week (Feb 25–Mar 3) without any tasks scheduled?
8. When is the next deadline due?
9. When does the next event take place?
10. How many things do you have scheduled for tomorrow (Feb 19)?
11. Who has requested from you text for the brochure before the end of the week (Feb 22)?
12. Are there any things to do related to a library?
13. Do you have any visits outside your school/workplace scheduled for the next week (Feb 25–Mar 1)?
14. What day of the next week (Feb 25–Mar 1) do you have scheduled a visit outside your school/workplace?
15. How many things do you have scheduled for today (Feb 18)?
16. How many events do you have to attend on Friday February 22?
17. What is the first day without any tasks scheduled?
18. Do you have any work-days in February without any tasks scheduled?
19. Do you have any tasks to do next Wednesday in the afternoon?
20. What time is the group meeting next Tuesday (Feb 26)?
21. Is the meeting next Monday (Feb 25) in the afternoon or in the morning?
22. How many tasks do you have scheduled for the next working week (Feb 25–Mar 1)?
23. What day is the meeting to discuss technology scheduled for?
24. When is the Book Club Meeting scheduled?
25. How many people will be attending the meeting on Tuesday Feb 26th?

26. What day and time is the next event scheduled for?
27. What is the meeting on Feb 25 about?
28. Who is giving the seminar on Feb 28?
29. Do you have any things planned for April?
30. Have you been invited to any parties in the next four weeks (-Mar 10)?
31. Can you schedule lunch on Tuesday Feb 26 at 12:30pm?
32. Do you have any coming weekends in February with nothing planned?
33. For how many weekends in March do you have anything planned?
34. What was requested from you by Chris?
35. Who asked that you do a guest lecture on New Technologies?
36. When are you going to give your guest lecture?
37. What do you need to do for the brochure?
38. Whom are you meeting on Tuesday Feb 26th?
39. Can you schedule more events on February 26?
40. Can you schedule lunch this Friday 12:30pm?
41. Can you schedule anything on Tue, Feb 26 at 3pm?
42. When is the SIGs Panel Session?
43. A friend calls you to set a dinner date for this Thu Feb 21 or Fri Feb 22. Are you free for dinner on Thu or Fri?
44. Can you schedule more things on Tue Mar 5 in the afternoon after 2pm?
45. Can you schedule more things on March 11 in the afternoon?
46. What kind of sports games do you have scheduled?
47. When do you have scheduled a sports game?
48. What membership do you need to renew?
49. Can one renew the membership in your Professional Association electronically?

References

Baddeley, A., 1986. Working Memory, Clarendon Press, Oxford.

Bälter, O., Sidner, C., 2000. Bifrost inbox organizer: giving users control over the inbox, Technical Report: Lotus Research TR 00-08.

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., 2003. Taking email to task: the design and evaluation of a task management centered email tool, in: Proceedings of the Conference on Human Factors in Computing System CHI'2003, ACM Press, New York, NY, pp. 345–352.

Carroll, J.B., 1974. Psychometric tests as cognitive tasks: a new 'structure of intellect'. Research Bulletin, 74.

Carroll, J.B., 1993. Human Cognitive Abilities, Cambridge University Press, Cambridge.

Cattel, R.B., 1971. Abilities: their Structure, Growth and Action, Houghton Mifflin, New York.

Chen, C., Rada, R., 1996. Interacting with hypertext: a meta-analysis of experimental studies. Human–Computer Interaction 11(2), 125–156.

Conway, M.A., 1990. Autobiographical Memory: an Introduction, Open University Press, Milton Keynes.

Cypher, A., 1986. The structure of users' activites. In: Norman, D., Draper, S.W. (Eds.), User Centered Systems Design: New Perspectives on Human–Computer Interaction, New Perspectives Associates, Hillsdale, pp. 243–263.

Czerwinski, M., Cutrell, E., Horvitz, E., 2000. Instant messaging and interruption: influence of task type on performance. In: Paris, C., Ozkan, N., Howard, S., Lu, S. (Eds.), Proceedings of OZCHI 2000: Interfacing Reality in the New Millennium, Academic Press, New York, pp. 356–361.

Czerwinski, M., Horvitz, E., Wilhite, S., 2004. A diary study of task switching and interruptions, in: Proceedings of ACM Conference on Human Factors in Computing Systems CHI', ACM Press, New York.

Dillon, A., Watson, C., 1996. User analysis in HCI: the historical lessons from individual differences research. International Journal of Human–Computer Studies 45(6), 619–637.

Ducheneaut, N., Bellotti, V., 2001. Email as habitat: an exploration of embedded PIM. Interactions 8(5), 30–38.ACM Press, New York, NY.

Egan, D., 1988. Individual differences in human–computer interaction. In: Helander, M., (Ed.), Handbook of Human–Computer Interaction, Elsevier, Amsterdam, pp. 543–568.

Ekstrom, R.B., et al., 1976. Manual for Kit of Factor Referenced Cognitive Tests, Education Testing Service, Princeton, NJ, USA.

Farnham, S., 2002. Personal Map: Automatically Modeling the User's Online Social Network. Paper presented at CSCW 2002 workshop: Redesigning Email for 21st Century.

Hutchins, E., 1995. Cognition in the Wild, MIT Press, Cambridge, MA.

Gruen, D., 1996. The role of external resources in the management of multiple activities. Unpublished PhD Thesis. UCSD, San Diego, CA, USA.

Gwizdka, J., 2001. Supporting prospective information in email, in: Extended Abstracts of ACM Conference on Human Factors in Computing Systems—CHI'2001, ACM Press, New York, NY, pp. 135–136.

Gwizdka, J., 2004. Cognitive abilities and email interaction: impacts of interface and task. Unpublished PhD Dissertation. University of Toronto.

Gwizdka, J., Whittaker, S., 2003. Redesigning Email for the 21st Century. Report from CSCW'02 Workshop, SIGCHI Bulletin, 2003, March/April, vol. 35. ACM Press, New York, NY.

Jackson, T.W., Dawson, R.J., Wilson, D., 2001. The cost of email interruption. The Journal of Systems and Information Technology 5, 81–92.

Jovicic, A., 2000. Implications for the design of email management software. Unpublished MSc Thesis, Department of Computer Science, University of Toronto.

Kaptelinin, V., 1996. Activity theory: implications for human–computer interaction. In: Nardi, B., (Ed.), Context and Consciousness: Activity Theory and Human–Computer Interaction, MIT Press, Cambridge, MA, pp. 353–360.

Kaptelinin, V., 2003. UMEA: translating interaction histories into project contexts, in: Proceedings of the Conference on Human Factors in Computing Systems CHI'2003, ACM Press, New York, NY, pp. 353–360.

Kline, P., 2000. The New Psychometrics, Routledge, London and Philadelphia.

Kuutti, K., 1996. Activity theory as a potential framework for human–computer interaction research. In: Nardi, B.A., (Ed.), Context and Consciousness: Activity Theory and Human–Computer Interaction, MIT Press, Cambridge, MA.

Lindsay, P., Norman, D.A., 1977. Human Information Processing, Academic Press, New York.

Lohse, G., 1997. The Role of Working Memory on Graphical Information Processing. *Behaviour & Information Technology*, 16, 297–308.

Mackay, W.E., 1988. More than just a communication system: diversity in the use of electronic mail, in: Proceedings of Conference on Computer-Supported Cooperative Work CSCW'1998, ACM Press, New York, NY, pp. 26–28.

Malone, T., 1983. How do people organize their desks? Implications for the design of office information systems, ACM Transactions on Information Systems, vol. 1. ACM Press, New York, NY, pp. 99–112.

McFarlane, D.C., Latorella, K.A., 2002. The scope and importance of human interruption in human–computer interaction design. Human–Computer Interaction 17, 1–61.

Miyata, Y., Norman, D., 1986. Psychological issues in support of multiple activities. In: Norman, D., Draper, S.W. (Eds.), User Centered Systems Design: New Perspectives on Human–Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, pp. 265–284.

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.

Modjeska, D., Chignell, M., 2003. Individual differences in exploration using desktop VR. Journal of the American Society for Information Science and Technology, 54(3), 216–228.

Murphy, K.R., Myors, B., 1998. Statistical Power Analysis. A Simple and General Model for Traditional and Model Hypothesis Tests, Lawrence Erlbaum Associates, Hillsdale.

Neuwirth, C.M., et al., 1998. Envisioning communication: task-tailorable representations of communication in asynchronous work, in: Proceedings of the Conference on Computer-Supported Cooperative Work, CSCW'1998, ACM Press, New York.

Oulasvirta, A., Saariluoma, P., 2004. Long-term working memory and interrupting messages in human–computer interaction. Behaviour and Information Technology 23, 53–64.

Robertson, G., Mackinlay, J., Card, S.K., 1991. Cone Trees: Animated 3D Visualizations of Hierarchical Information, ACM Press, pp. 189–194.

Rohall, S.L., Gruen, D., Moody, P., Kellerman, S., 2001. Email vsualizations to aid communications, in: Proceedings of IEEE Symposium on Information Visualization InfoVis'2001, IEEE Press, New York, pp. 12–15.

Rohall, S.L., Gruen, D., 2002. ReMail: A reinvented email prototype (Technical Rep. no. 02–13) IBM.

Sein, M.K., Bostrom, R.P., Olfman, L., Davis, S.D., 1993. Visualization ability as a predictor of user learning success. International Journal of Man–Machine Studies 39, 599–620.

Stern, M.K., 2003. Identifying and understanding dates and times in email, Technical Report. (Rep. No. TR2003-11).

Sudarsky S. & Hjelsvold R., 2002. Visualizing electronic mail. In Proceedings of Visualisation Conference IV'2002, NY, NY: ACM Press, pp. 3–9.

Venolia, G.D. & Neustaedter, C., 2003. Understanding sequence and reply relationships whitin email conversations: a mixed-model visualization. In Conference Proceedings on Human Factors in Computing Systems CHI'2003, NY, NY: ACM Press, pp. 361–368.

Westerman, S.J., 1993. Individual differences in human–computer interaction. Unpublished. PhD dissertation, Aston University England.

Westerman, S.J., 1995. Computerised information retrieval: Individual differences in the use of spatial and non-spatial information. Perceptual and Motor Skills 81, 771–786.

Westerman, S.J., Cribbin, T., 2000a. Mapping semantic information in virtual space: dimensions, variance, and individual differences. International Journal of Human–Computer Studies 53(5), 765–788.

Westerman, S.J., Cribbin, T., 2000b. Cognitive ability and information retrieval: when less is more. Virtual Reality 5(1), 1–7.

Whittaker, S., Sidner, C., 1996. Email overload: exploring personal information management of email, in: Proceedings of the Conference on Human Factors in Computing Systems CHI'1996, ACM Press, New York, pp. 276–283.

Whittaker, S., Terveen, L., Nardi, B., 2000. Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. Human–Computer Interaction 15, 75–106.

Whittaker, S., Jones, Q., Terveen, L., 2002. Managing long term conversations: conversation and contact management, in: Proceedings of the 35st Hawaii International Conference on System Sciences—HICSS-35.

Woods, D.D., 1984. Visual momentum: a concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*. Academic Press, 21(3), 229–244.

Yiu, K., Baecker, R.M., Silver, N., Long, B., 1997. A time-based interface for electronic mail and task management, in: Proceedings of HCI International '1997, vol. 2. Elsevier, Amsterdam, pp. 19–22.

Zhang, J., Norman, D.A., 1994. Representations in distributed cognitive tasks. Cognitive Science 18, 87–122.

Zhang H., Salvendy, G., 2001. The implication of visualization ability and structure preview design for web information search tasks. International Journal of Human–Computer Interaction 13, 75–95 .