Improving shift handovers between medical residents : the effect of information access cost, metacognition and media affordances

Yang, Xi

2014

Yang, X. (2014). Improving shift handovers between medical residents : the effect of information access cost, metacognition and media affordances. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/61053

https://doi.org/10.32657/10356/61053

IMPROVING SHIFT HANDOVERS BETWEEN MEDICAL RESIDENTS: THE EFFECT OF INFORMATION ACCESS COST, METACOGNITION AND MEDIA AFFORDANCES

YANG XI

School of Mechanical and Aerospace Engineering

A thesis submitted to Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2014

Acknowledgement

This thesis could not have been written without my advisors, Professors Leong Kah Fai, Christopher D. Wickens, Park Taezoon and Martin G. Helander. They guided, encouraged and challenged me along the way, never accepting less than my best efforts. I am deeply indebted to them. Also, I want to extend my appreciation to Dr. Halimathun Khalid for her insights on how to make this research more robust.

I would like to express my gratitude to Drs. Kewin Siah, Liesel Fong and Sophia Ang from National University Hospital, Singapore, and Dr. Yoel Donchin from Hadassah Hebrew University Hospital, Israel, for being excellent Subject Matter Experts. Special thanks go to Yoel who introduced me to patient safety, and Kewin who was the clinical champion for this research. I also wish to thank Drs. Yin Shanqing and Lim Tiek Whai from Changi General Hospital, Singapore, for their support throughout the data collection.

All the students from the Human Factors lab supported me, especially my fellow colleagues and friends Benedict, Ranieri and Syahid. I am thankful for their constructive input to the research and all the funs we had in the lab.

Last but not least, I would like to thank my parents for their endless love. They have always encouraged me to pursue my curiosity and this thesis would have been impossible without them.

Table of contents

List of figures

List of tables

Summary

The continuity of patient care and limited resident working hours routinize the handover of patients between doctors. Despite of the pivotal role to ensure quality of patient care, clinical handovers were found to be vulnerable episodes contributing to near misses, adverse events and waste of healthcare resources. Handovers can be conceptualized as a three-phase process, namely pre-handover, handover communication, and post-handover. Existing research, however, is largely limited to the communication phase, especially to the study of communication content. Little effort was made to study the information retrieval strategy during pre-handover and the effect of media during handover communication. This thesis therefore aims to propose a suitable framework for modeling the factors influencing the performance of pre-handover and handover communication; and to study experimentally the effects of information access cost, overconfidence bias and media affordances.

This thesis includes an extensive review of clinical shift handovers, information retrieval strategy and media affordance, as well as four empirical studies to investigate the proposed research models. Study 1 explored medical residents' handover practices in a university hospital. Study 2 adopted a laboratory setting to examine the underlying factors affecting the information retrieval strategy. Study 3 investigated medical residents' information retrieval strategy in pre-handover. Study 4 examined the effect of media affordances on medical residents' handover communication.

With regard to the pre-handover phase, the results of quantitative and qualitative analyses across the studies showed that besides the well-documented factor - information access cost, the perceived accuracy of memory is another important determinant influencing the information retrieval strategy. Moreover, overconfidence affected medical residents' pre-handover performance.

With regard to the handover communication phase, the results of quantitative and qualitative analyses across the studies showed that reviewability (14% increase) and visual access to object of reference (12% increase) strongly supported the handover receivers' memory retention of important patient information, and interactivity supported both memory retention (24% increase) and case comprehension (26% increase). Furthermore, the results provided contradicting evidence to the long-held recommendation that face to face handover is preferred to mediated communication. Being able to see other was found to affect neither memory retention nor case comprehension.

This research is of both theoretical and practical significance. Theoretically, the research proposed two research models to describe various factors influencing the performance of pre-handover and handover communication. In particular, at the pre-handover phase, this research modeled the relationship between information access cost, overconfidence bias and medical residents' information retrieval strategy. At the handover communication phase, this research modeled the relationship between the affordances of interactivity, reviewability, visual access of object of reference, visibility and medical residents' handover communication performance. Moreover, this research provided empirical evidence for the causal relationships. Alongside the theoretical significance is a very practical concern. Singapore is aiming to become a leading medical hub in Asia and hospitals in Singapore are developing standard handover protocols to enhance patient safety. Results from this research provided practical implications for the development of handover tools and protocols.

1. Introduction

1.1 Background

In 2009, The New York Times featured a heart-breaking story in the article "When patient handoffs go terribly wrong (Chen, 2009)". In the story, a two-year old boy Joey went through a surgical operation to remove a mass on the left side of his face. It was a difficult operation and completed by one senior surgeon and one covering surgeon. Two days after the operation, Joey's parents found his left cheek remained frozen and he would never smile symmetrically because of the accidental cut of facials nerve underneath the mass. Although the exact time when the nerves were damaged was unknown, the poor handover between the two surgeons was suggested to be the major reason which might have caused the accident. The story created a huge impact in the medical community back then and its ripples can still be felt today.

The continuity of patient care and limited resident working hours routinize the transfer of patients between healthcare professionals. This transfer process is called handover and it is defined as "the transfer of role and responsibility from one person to another in a physical or mental process (Solet, Norvell, Rutan, & Frankel, 2005)". The primary objective of a handover is to provide accurate information about a patient's state and plan of care (Patterson, Roth, Woods, Chow, & Gomes, 2004) .

Clinical handovers are essential in healthcare communication. In Australian hospitals alone, there are approximately 7,068,000 handovers annually (ACQSHC, 2009). Despite of the pivotal role to ensure quality of patient care, clinical handovers have historically been an unstructured process and have not received much attention until recently. Studies have shown that clinical handovers are high risk scenarios for patient safety. Poor handovers have been associated with delayed treatment, longer patient stays, more laboratory tests, clinical near misses and errors, and self-reported harm to patients (Horwitz, Moin, Krumholz, Wang, & Bradley, 2008; Kitch et al., 2008; Singh, Thomas, Petersen, & Studdert, 2007) .

1.2 Research motivation

A handover can be conceptualized as a three-phase process, consisting of pre-handover, handover communication, and post-handover (Kerr, 2002; Raduma-Tomàs, Flin, Yule, & Close, 2010). The main objective of the pre-handover phase is to ensure that an outgoing professional is ready to transfer patients to an incoming professional. It involves activities of collecting handover information and preparing handover notes. During the handover phase, the two professionals communicate with each other using various methods. In the post-handover phase, the incoming professional prioritizes handover tasks, plans his/her duty and performs necessary actions. Since accurate and adequate handover preparation and communication are prerequisites for proper post-handover performance, this research focuses on the first two phases.

Studies on the pre-handover phase are scarce. Results from the limited literature showed that outgoing doctors tended to use memory-intensive strategies to retrieve patient information, by which they relied upon their memory rather than turned to written documentations such as patient medical records. Many residents handed over verbally “wherever and whenever” they could find the on-call doctor without proper preparation of handover notes (Vidyarthi, Arora, Schnipper, Wall, & Wachter, 2006). In addition, very few information tools were used by the outgoing doctors during the pre-handover phase, with the exception of a recent study by

Hilligoss and Zheng (2013). Outgoing doctors mainly wrote down minimal information and much additional information would then be recalled and presented during the handover communication (Alem, Joseph, Kethers, Steele, & Wilkinson, 2008; Yang et al., 2011). Such memory-intensive strategies impose potential threats to patient safety, since human memory is vulnerable to errors (Wickens, Hollands, Parasuraman, & Banbury, 2013). However, there were no studies addressing the reason why doctors prefer memory-intensive strategies, and quantifying the potential negative impacts on patient safety.

A majority of existing studies on the communication phase focused on the information content being communicated (Cohen & Hilligoss, 2010). The general approach is to quantify the information communicated and information omitted. These studies explored the information coverage in handovers and offered important design guidelines on handover content standardization. One such example is the proposition of SBAR (situation, background, assessment and recommendation) (Leonard, Graham, & Bonacum, 2004). However, communication content is not the entire story, as communication in social systems relies on communication media to transfer information. Communication media determine the channels through which information can be delivered and received, and hence affect communication processes and outcomes (Welch, Cheung, Apker, & Patterson, 2013). Moreover, the advance in information communication technology (ICT) is at an unprecedented speed providing the healthcare professionals with new media to transfer information across time and location. Therefore, studying communication media is equally important if researchers aim to fully understand clinical handovers and to enhance patient safety. There are only a limited number of studies addressing how communication media affect handover processes and results. They were largely based on observations in hospital settings. Although ethnographic observations yield valuable insights on how handovers happen in real settings, it is difficult to determine any causal relationship by these studies. In general, these studies suggested that direct communication (e.g. face to face handoffs, handoffs at bedside) was preferred to mediated communication (e.g. handoffs via telephone, email) in patient handovers. Besides observational studies, only a few experimental studies explored the advantages and disadvantages of several handover methods (Bhabra, Mackeith, Monteiro, & Pothier, 2007; Pothier, Monteiro, Mooktiar, & Shaw, 2005). However, these studies were not based on sound theoretical foundations. Thus their results had limited power of generalization, though they might be of high value to the particular settings.

1.3 Research objectives and significance

To address the identified research gaps, this research aims to propose research models for delineating the factors influencing the performance of pre-handover preparation and handover communication; and to study experimentally the effects of information access cost, overconfidence bias on medical residents' information retrieval strategy during pre-handover phase and the effects of media affordances on handover communication performance.

This research is of both theoretical and practical significance. Theoretically, the research proposed two research models to delineate the various factors influencing the performance of pre-handover preparation and handover communication. In particular, at the pre-handover phase, this research modeled the relationships between information access cost, overconfidence bias and medical residents' information retrieval strategy and performance. At the handover communication phase, this research modeled the relationships between the affordance of interactivity, reviewability, visual access of object of reference, visibility and medical residents' handover communication performance. Moreover, this research provided the empirical evidence for the causal relationships described in the two models. Alongside the theoretical significance is a very practical concern. Singapore is aiming to become a leading medical hub in Asia and hospitals in Singapore are developing standard handover protocols to enhance patient safety. Results from this research provided practical implications for the development of handover tools and protocols.

1.4 Research scope

This research limits its scope to the pre-handover phase and the handoff communication phase. In addition, it addresses handovers between medical residents during shift changes. It did not address handovers between groups of doctors, or between senior clinicians, nor other types of handovers such as interdisciplinary transfer of patients.

1.5 Structure of the thesis

Figure 1-1 shows the structure of the thesis. Chapter 1 introduces the background of this research, followed by the research motivation, research objectives, research scope and structure of the report. Chapter 2 presents a literature review of theoretical frameworks and past studies related to the analysis of the problem domain. Chapter 3 summarizes the research gaps identified from the literature review and presents the research framework and hypotheses. Chapter 4 presents the results from an exploratory study conducted in a major university hospital in Singapore, and discusses how the results support the research framework. Chapter 5 presents Study 2 examining the factors influencing information retrieval behavior in general, followed by Study 3 in Chapter 6 applying the same research model to investigate medical residents' information retrieval behavior and performance during pre-handover. Chapter 7 presents Study 4 examining the effect of communication media on handover quality. Chapter 8 concludes the dissertation.

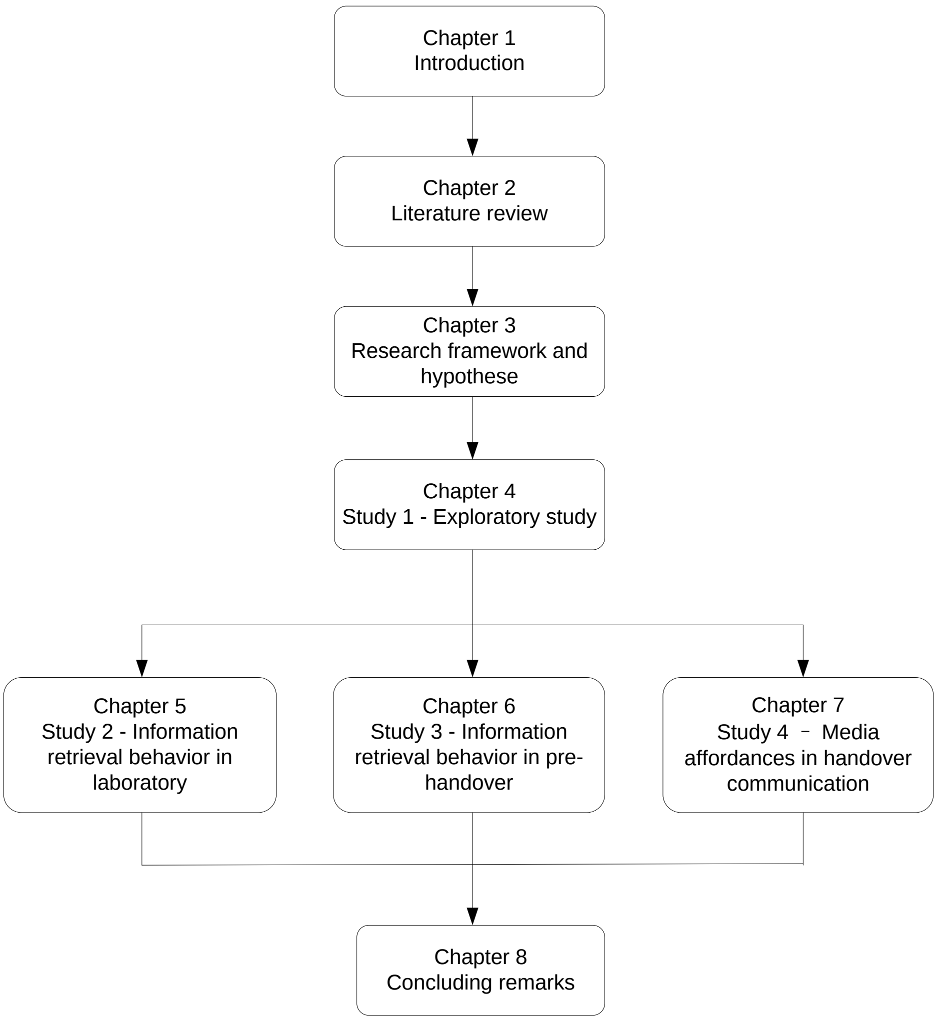


Figure 1-1 Structure of thesis

2. Literature review

This chapter begins with a review of research on pre-handover and handover communication in healthcare. The subsequent two sections cover theories and literature pertaining to doctors' information retrieval behavior during pre-handover and the study of communication media during handover communication. The last section summarizes the research gaps.

2.1 Pre-handover and handover communication in hospitals

2.1.1 Information preparation during pre-handover

As stated in the introduction, the objective of the pre-handover phase is to ensure that an outgoing professional is ready to hand over to an incoming professional; and the main activities of the pre-handover phase are to collect handover information and prepare handover notes. There is a very limited amount of literature on healthcare practitioners' pre-handover activities, which is reviewed below.

Lawrence, Tomolo, Garlisi and Aron (2008) posited that the pre-handover phase could easily be lost or truncated in a chaotic environment such as in hospitals, resulting in compromised or lost strategies related to the phase. For example, doctors may lose track of time and suddenly realize that it is time to hand over patients, and hence deliver what they can recall from memory. This claim was supported by three qualitative studies. Vidyarthi et al. (2006) surveyed residents at three hospitals and reported no uniformity in how they handed over to cross-covering doctors. Many residents handed over verbally "wherever and whenever" they could find the cross-covering doctor without proper preparation of handover notes. Due to the lack of preparation, it is likely that residents depend on their own memory when transferring patient information.

Alem et al. (2008) studied information tools used by outgoing doctors during handover in a general medical ward and an Emergency Department (ED). They found that in both facilities very few information tools were used. Outgoing doctors mainly wrote down small bits of patient information on patient lists (mainly in ED) or on blank sheets of paper (mainly in the general medical ward). Such information in the notes was often used as a ‘memory trigger’ and much additional information would be recalled and presented during handovers. Moreover, medical records were seldom used during handovers in either ED or in the general medical ward. Doctors claimed that they often remembered patients who had been under their care. However, as the authors stated, there was no guarantee that a doctor would remember correctly the patient information.

Yang et al. (2011) interviewed residents in a university hospital in Singapore about their handover practices using a three-phase prospective. The results showed that there was no structured protocol implemented for the pre-handover phase. Outgoing doctors handed over patients to on-call doctors at their own discretion. In addition, when retrieving patient information outgoing doctors tended to depend on their memory rather than referring to various kinds of written documentation such as patient charts.

Contrary to doctors’ ad-hoc practices during pre-handover, a study from Kerr (2002) on nurse handover practices reported a fairly structured pre-handover process. The author observed the participating nurses from different wards and documented a variety of strategies that nurses adopted in the pre-handover phase. Nurses of the

Oncology/Hematology wards were concerned with updating the official documents, such as care plans and the ward diary. One or two nurses gathered information from colleagues, official documents and their personal notes. On the ENT/Plastic/Dental ward, more outgoing nurses were involved in pre-handover activities. Each nurse left the ward with the official communication book and their own notes, in order to record information about the patients they had been caring for. The official documentation was passed among the off-going staff until a full report had been recorded.

Besides the study of handover senders' information collection during pre-handover, Hilligoss & Zheng (2013) conducted an ethnographic study to examine how handover receivers (inpatient clinicians) used electronic health records (EHRs) to in preparation for handovers between ED clinicians and inpatient clinicians. The authors reported three advantages of adopting such practices: getting an overview of the patient, preparing for handover and subsequent care, and protecting against errors.

Although limited, these studies suggest that there is a variety of practices that healthcare practitioners follow in the pre-handover phase. Nurses adopted a fairly structured process by collecting patient information from various written documents and cross-examining the information accuracy and adequacy with their colleagues. Doctors, on the contrary, followed an ad-hoc process and were more likely to adopt on memory-intensive strategies. The adoption of memory-intensive strategies may impose potential threats to patient safety, as human memory is vulnerable to errors (Wickens et al., 2013). However, the underlying reasons for such strategies remain unknown.

2.1.2 Content and media in handover communication

Studies on handover communication phase are mainly concerned with handover content and media. A majority of existing studies on the communication phase focused on the information content being communicated. The general approach is to quantify the information communicated and information omitted. These studies explored the information coverage in handovers and offered important design guidelines on handover content standardization. However, communication content is not the entire story, as communication in social systems relies on communication media to transfer information. Communication media determine the channels through which information can be delivered and received, and hence affect communication processes and outcomes (Welch et al., 2013). Therefore, studying communication media is equally important if researchers aim to fully understand clinical handovers and to enhance patient safety. There are only a limited number of qualitative studies addressing how communication media affect handover processes and results. In general, the studies suggested that direct communication (e.g. face to face handovers, handovers at bedside) is preferred to mediated communication (e.g. handovers via telephone, email) in patient handovers.

2.1.2.1 Handover content

Researchers have been striving to understand which information should be included in a handover. On one hand, de-emphasizing or omitting crucial information that is critical to manage the patients can lead to adverse events and near misses. On the other hand, communicating too much information or irrelevant information may burden the receivers, resulting in less attention paid and less information remembered, and may ake up time that could be spent on other tasks.

A number of studies have addressed the omission of crucial information during handovers. In two qualitative studies from Sutcliffe, Lewton, and Rosenthal (2004) and Arora, Johnson, Lovinger, Humphrey, and Meltzer (2005), residents expressed their concerns that the attending physicians/fellow residents provided far too little information to the residents who would be caring for patients during on-call periods.

Ye, McD Taylor, Knott, Dent and MacBean (2007) investigated quantitatively the impact of content omission. They analyzed the handovers of 914 patients and found that information was perceived as lacking in 109 (15.4%) handovers. The perceived quality was significantly higher when all required information was handed over. As a result of the inadequate information transfer, the doctor and the patient were affected adversely in 62 (8.8%) and 33 (4.7%) cases, respectively.

Borowitz, Waggoner-Fountain, Bass and Sledd (2008) studied the residents' sign-out at two general pediatric wards over 98 days. After each night of duty, the participants were asked to complete a post-call survey, focusing on the sign-out information they received previously. In 49 out of 158 (31%) surveys, residents indicated something happened while they were on call for which they were not well prepared. In 40 of the 49 cases, they indicated that the transferred information was inadequate, and the missing information would have been useful.

Lamond (2000) compared the content of the shift report and notes information on the medical and surgical wards in National Health Service (NHS) hospitals. Five consecutive reports on each ward were audio-taped, and the medical notes and nursing documentation (care plans, all observation charts, medication chart, etc.) for 15 patients on each ward examined, giving a total sample of 60 patients. Each of these sources of data was then subjected to content analysis. The frequency data indicated that almost all types of information were contained in the notes and charts of patients more frequently than in the report, whereas global judgments made about the patients' condition and psychological state/ personality were mentioned more in reports than in patients' notes. Moreover, the information available from report appears to follow a certain profile, with some general information and information about the patients' condition being given, but often summarized in the form of a global judgment. The concept of global judgment is also described in two handover training programs by Patterson, Weiss, Woods, Mount-Campbell, and Rayo (2013) and DeVoge et al. (2009).

Compared to the considerable amount of research on ‘communicating too little’, there is little research on the transfer of irrelevant or unnecessary information during handovers in healthcare. One exception is the critical incident analysis conducted by Arora et al. (2005). Besides the frequently cited ‘content omission’ as one of the causes of poor handovers, transfer of irrelevant information, was also mentioned by the participants. It was illustrated by the statement from one participant – “The written sign-out has a different role than just a capture of a patient. We use it to keep track of everything but sometimes that makes it difficult to know what is going on right now since there is so much extra information on it.” In another study, Carroll (2012) documented the outgoing nurses adaptive strategies in handover based on the incoming nurses’ knowledge of the patient. They found that the handover reports were shorter if the incoming nurses already know a patient as the outgoing nurses perceived a lower necessity to discuss every single details of the patient with the incoming nurses. Moreover, Wears (2012) questioned the data transfer framing of handover, that "more items are always better than fewer". The author argued that it is practically impossible to transfer complete information learnt about a patient during the previous shift and the value of a good handover is that it is not complete."

In summary, studies of doctors' handovers have demonstrated that content omission is not rare in real practice, but there is little research studying the impact of information overload on handover outcomes.

2.1.2.2 Handover media

Compared to the amount of research on handover content, there are only a small number of studies on handover method/media. Research on handover methods can be broadly categorized into three themes according to their focus. The first theme is focused on the extent to which a handover method is able to provide interactive channels in handover communication. The Joint Commission has recommended that a proper handover protocol should include an opportunity for the participants to ask questions and clarify doubts (2006). In addition, Patterson and Wears (2010) discussed "Resilience" as a framing of handover functions. The central idea of "Resilience" lies in the opportunity for the incoming provider to ask clarification and error-detection questions. A few observational studies in medical handovers have supported the conclusion that an interactive medium results in better handoff results (Horwitz et al., 2009; Obstfelder & Moen, 2006; Randell, Wilson, & Woodward, 2011). Obstfelder and Moen (2006) explored the consequences of introducing an electronic patient record (EPR) system to support community health services in a Norwegian municipality. They found that staff were reluctant to use the EPR system during handovers. Further exploration of the interaction with the EPR showed that it eliminated the oral conversation part which permits interaction between off-going and on-coming parties. Along the same line, Horwitz et al. (2009) evaluated the effect of changing the form of communication between the handover sender - emergency department (ED) doctor and the receiver – general medical ward doctor at admission. An intervention was introduced to replace synchronous (telephone or face to face) communication by asynchronous communication (voicemail). Before the intervention, both doctors from ED and the general medical ward thought the voicemail system would be a bad idea. But after experiencing the new system, the ED doctor was generally more positive. Both doctors from ED and general medicine reported reduced interaction due to the asynchronous communication. In 30% of the cases, the communications receiver in this study, completely ignored the new system. They never listened to the voicemail, but rather started their assessment of a new admission with no input from the emergency physician at all. In another study, Randell, Wilson and Woodward (2011) conducted a multi-site case study to investigate nursing and medical face-to-face verbal handover and reported that through verbal communication the incoming healthcare professionals were able to seek further information and clarify doubts easily.

The second theme of research is related to the visual information that handover senders and receivers view during handover communication. It has to be noted here that there are two kinds of visual information – the view of the communication parties, in this case, the outgoing and incoming clinicians or the view of the task space, in this case, the patient and the patient-related information (Watts & Monk, 1996). Solet et al. (2005) observed the patient handover process at Indiana

University School of Medicine's internal medicine residency and based on the observation, argued that direct communication (e.g. face to face handovers, handovers at bedside) is superior to mediated communication (e.g. handovers via telephone, email) in patient handovers. The reason was that direct communication could afford a full spectrum of communication channels including facial expression, gesture, and eye contact and so on. Patterson and Wears (2010) described "transfer of responsibility and authority" as one framing of handover functions. The focus of the framing is to ensure that patients are assigned to a provider successfully, that a provider has accepted the transfer, and that other staff are aware of who is responsible for a patient. Being able to see communication parties may affect positively the transfer of responsibility and authority. In another study, Anderson and Mangino (2006) implemented bedside nurse handovers at one adult acute care unit and reported positive results from conducting nurses' handovers at bedside, including financial savings and increased patient and staff satisfaction. Staff members commented that beside handovers enabled them to prioritize their shift work better because they had visualized all of their patients within the first 20 to 30 minutes of their shift. Increase in staff satisfaction can be explained by the enriched communication medium, in particular the affordance of viewing the patients besides verbal handovers. In another case study, Randell et al. (2011) described a case of error resolution by outgoing and incoming nurses. By reviewing the task space – patient medical records, the outgoing and incoming nurses clarified a medial test that the patient would be taken during the shift.

The third theme is related to the inclusion of written notes in handover communication. Pothier et al. (2005) have conducted the first experimental study to investigate the effect of handover methods. They utilized a simulated setting to study the effect of different handover methods on nurses' information retention during handovers. Three types of methods were tested, including verbal handover, verbal handover with note-taking, verbal handover with typed handout. Five nurses participated in their study, handing off 12 fictional patients from one to the next consecutively. The results showed that pure verbal handover led to a complete loss of data after three handover cycles. The verbal handover with notes-taking style resulted in 69% of data loss after five cycles and the verbal handover with typed handout demonstrated minimal loss of data. All the differences were statistically significant. In a following study, Bhabra et al. (2007) replicated the same setting to study handovers between doctors, with 5 senior house officers being the subjects. The results revealed similar patterns as the nurses' handovers. In the pure verbal handovers, only 2 out of 80 (2.5%) data points were retained after 5 handover cycles, while 85.5% of information was retained in the note-taking case and 98.75% in the typed handout group. These two studies clearly demonstrated the difference in efficacies of the handover styles. Pure verbal handover is not an effective handover method, while the typed handout is the most effective way of retaining information. Also, the authors recommend a combination of verbal and written communication, as this provides multiple channels for information to be exchanged. Despite the advantage of having a typed handout during handover, Philibert (2009) identified a potential vulnerability of using handouts, because it could result in a ''go down the list and report all information'' paradigm, with irrelevant information and less important information included in the handouts.

In summary, there are only a limited number of studies addressing how communication media affect handover processes and results. They were largely based on field observations in real settings. Although ethnographic observations yield valuable insights, they were potentially affected by other confounding factors, making it difficult to fully disentangle the effect of the communication media from the confounding factors. Only a few experimental studies explored the advantages and disadvantages of several handover methods (Bhabra et al., 2007; Pothier et al., 2005). However, these studies were subject to two major shortcomings. First, the experiments were based on small sample size, thus their results may have limited power of generalization. Second, the experiments were of limited scope. Specifically, they only varied the handover receivers' ability of having written information. Therefore, more research is needed to uncover the relationship between various handover methods to handover quality.

2.2 Information retrieval during pre-handover

This section reviews four classes of literature relevant to the information retrieval behavior during pre-handover. First, it reviews the role of effort in choice. There is a great deal of literature here, but it does not address the specifics of information access effort. Instead much of it reviews the role of effort on the choice to behave safely; the effort involved in the choice to continue a visual search task, or the effort in the choice of one decision strategy over another. Second, it considers more specifically, the nature of information access effort (e.g., as required to retrieve written information on a patient). It covers the literature bearing on this construct, and its role in determining information retrieval strategies even as it does not specifically describe the tradeoff with perceived accuracy, for example, the perceived accuracy of remembering a patient's resuscitation status. Third, this section considers studies that have examined perceived accuracy, particularly as manifest in overconfidence in memory. These studies document the circumstances when people anticipate that their accuracy is (or will be) greater than it actually is; and what drives this overconfidence. Fourth, it covers the rare literature that has actually examined tradeoffs between effort (or anticipated effort) and anticipated accuracy, in a way that is compatible with the simple decision model presented above (Fennema & Kleinmuntz, 1995; Payne, Bettman, & Johnson, 1988). However this is primarily done in the context of the choice between decision strategies (e.g., low effort heuristics versus high effort algorithms; rather than the choice between information retrieval strategies.

2.2.1 Effort as a concept

Effort has been the focus of numerous studies in psychology (Newell, 1990) and decision theory (Kahneman, Slovic, & Tversky, 1982). According to Kahneman (1973), effort can be conceived from two perspectives, relating to the person or to the task. A person can invest different amount of effort at his own discretion while a task can demand different amount of effort to be completed (Wickens & MaCarley, 2007).

A consistent finding related to the person concept of effort is that humans have limited physical and mental resources and thus are inherently "effort conserving" when performing a task (Wickens & MaCarley, 2007). Examples of effort conservation include the 'edge effect' in visual search - participants reduce visual scanning of peripheral areas of a screen due to the increased effort of doing so (Parasuraman, 1986; Stelter & Wickens, 2006), and the "cost of compliance" in the study of unsafe behaviors - participants are less inclined to wear safety harness if the effort wearing it is high (Wickens, Lee, Liu, & Becker, 2004; Wogalter & Laughery, 2006)

Besides the person concept of effort, Kahneman (1973) posited that the availability of mental resources is mobilized through increasing arousal of the autonomic nervous system as it responds to increasingly difficult tasks. A difficult task can stimulate a person to invest more effort, whereas an easy task will not. The task concept of effort was further illustrated by the performance resource function (PRF) (Norman & Bobrow, 1975). According to the PRF, a task is classified by its relationship between resources and data. A 'resource limited' task is one for which providing more resources will always improve performance. A 'data limited' task is one for which performance is limited by the quality of data, rather than the amount of resources.

The task concept of effort is widely used in research on dual-task performance where an optimal allocation of attention is often required (Wickens & MaCarley, 2007). The well documented proximity compatibility principle was is another example illustrating the concept of task effort (Wickens & Carswell, 1995). An information reading task could be either integrated, where information from various sources are needed for a judgment, or the task could be stand-alone, where information from a single source is needed. The greater the integration requirement for a task, the greater is the total effort demand from working memory.

2.2.2 Information access effort

A specific area of research on effort is concerned with information retrieval effort. Ballard, Hayhoe and Peiz (1995) developed the Blocks World Task (BWT) paradigm to investigate the relationship between information retrieval effort and people's interactive strategies. In the task, subjects were asked to replicate a pattern of colored blocks. The display was divided into three areas: model, resource and workspace. The model area contained the block pattern to be copied; the resource area contained blocks to be selected; and the workspace was the area where subjects replicated the model (Figure 2-1).

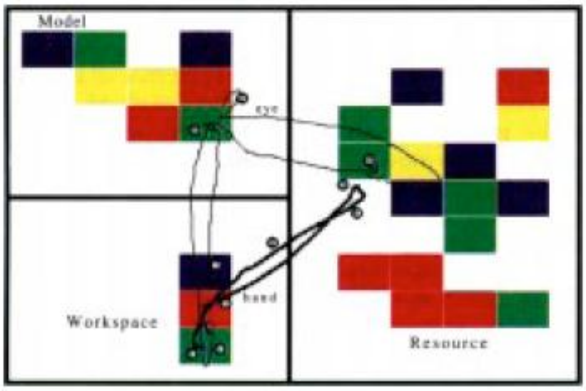


Figure 2-1 Display of model (top left), resource (right) and workspace (bottom left) (Source: Ballard, Hayhoe, Pook, & Rao, 1997)

The authors analyzed the strategies subjects used when performing the task and found that at the beginning of the task, the memoryless 'Model-pickup-model-drop' (MPMD) strategy was used most frequently. As the task proceeded, however, the memory intensive strategy 'Pickup-drop' (PD) was used increasingly and outweighed the other strategies at the end of the task. In a following study, Ballard, Hayhoe, Pook and Rao (1997) quantified the time needed for the different strategies, as shown in Table **2-1**. Using the memoryless MPMD strategy it took 3 seconds to place a block and the memory intensive strategy PD 1.5 seconds.

Table 2-1 Estimates of time taken to place a block for different strategies (Source: Ballard, Hayhoe, Pook, & Rao, 1997)

Strategy	Time (Sec)	Memory Items
MPMD	3	
PMD	2.5	color
MPD	2.0	offset
PD	1.5	color and offset

The changes in strategy can be explained as follows: at the beginning of the task, memorizing both the color and the location of a block was difficult and thus subjects adopted the memoryless, low-effort MPMD strategy. As the subjects proceeded with the task, they became more familiar with the color and the location of a block and hence the effort of retrieving the information of that block from memory decreased gradually. Therefore, the subject adopted the memory intensive PD strategy.

Gray and Fu (2001) proposed a "soft constraint" hypothesis to investigate the differences in retrieval behavior between information in the world and information in the head. They stated that there are two kinds of memories: internal memory and external memory, and effort is needed to retrieve information from the two kinds of memories. Internal effort (i.e., knowledge in the head, or KIH) refers to the effort used in encoding, storing and retrieval an item from memory. External effort (KIW) refers the effort spent searching the environment and the effort of retrieve an item at a known location. Furthermore, the soft constraint hypothesis claims that the selection of a high-level interactive behavior depends on the cost-benefit tradeoffs between the effectiveness and efficiency.

Despite the proposition of the two concepts (effectiveness and efficiency) a series of follow-up experiments focused solely on testing the effect of information access effort on the selection of interactive behaviors (Gray & Fu, 2001, 2004; Gray, Sims, Fu, & Schoelles, 2006). The first experimental setting was to program a VCR to record a television show. In order to complete the task, participants had to access 5 parameters including show name, show time, day of week, channel and program number. In addition, participants had to program a show correctly for two successive trials.

Internal effort and external effort needed to finish the task were manipulated through three scenarios: Free access condition – information was visible on the screen and the user had free access to information via an eye movement (low IAE for KIW); Gray box condition – desired information was covered by gray box and the user had to move the mouse and click on the box to view information (higher IAE for KIW); and memory test condition - desired information was covered as in gray box condition but the required information was well-learned before task begins (low effort for KIH). Temporal costs of memory retrieval and perceptual-motor access were estimated by ACT-R 5.0 model (J. R. Anderson, 1990) and CPM-GOMS model (Gray, John, & Atwood, 1993), as shown in Table **2-2**.

Access to knowledge in the world was used as the process measure and trials-to-criterion (number of trials a participant completes until two successive trials) and goal suspension (number of potential access errors) were used as the performance measure.

Table 2-2 Estimates of perceptual-motor and memory retrieval effort by condition (Source: Gray & Fu, 2004)

Condition	Perceptual-motor access	Memory retrieval
Free-access	500	500-1000 (weak)
Gray-box	1000-1500	500-1000 (weak)
Memory-test	1000-1500	100-300 (strong)

The soft constraints hypothesis made several predictions: The greatest amount of information window access would be made for free access condition, followed by gray-box condition, followed by memory test condition. The most number of trials-to-criterions and goal suspension would be made for gray-box condition, followed by free-access condition, followed the memory test condition. Results showed that the average information access was 1.31 checks per show at gray-box condition and 0.05 at memory-test condition. In addition, the free-access group was 4.3 times more likely to access information than the gray-box group. Also, the results showed a significant effect on trials-to-criterion and a marginal significant effort on goal suspension.

In a later study, Gray et al. (2006) used different methods to manipulate the IAE (IAC) and conducted three new experiments. The experimental setting was similar to Ballard’s study, except that a gray window covered the model, resource and workspace windows (modified BWT task). The costs of uncovering the target window were manipulated differently in the three experiments. In experiment 1, three levels of access cost were varied. In the low-cost condition, the target window opened when the subjects pressed a control key and stayed open until the control key was released. In the middle-cost condition, the target window opened when the cursor entered and closed when the cursor left. In the high-cost condition, a one second lockout time was imposed when the cursor entered the target window.

In experiment 2, the subjects opened the target window by moving the cursor to a button located at the center of the target window. The access cost was manipulated by using different sizes of the button – large size for low-cost condition, middle size for middle-cost condition and small size for high-cost condition. According to Fitts' law, the Fitts Index of Difficulty was expressed as

$$\mathrm{ID} = \log_2\left(\frac{D}{W} + 1\right),$$

where D is the distance to the target and W is the width of the target. Thus different size of the button created different access difficulty and hence access cost.

In experiment 3, the lockout time to uncover the target window was changed. Data for four lockout conditions, 0, 200, 1600, 3200 milliseconds was collected.

Across the three experiments, the number of target window accesses, the duration of first look and the number of blocks correctly placed following the first look were measured. The results revealed that as information access cost increased, the number of target window accesses decreased, duration of first look increased and the number of blocks correctly placed increased, suggesting a consistent shift from a memory-less strategy to a memory dependent strategy.

Utilizing the soft constraint hypothesis, Morgan, Patrick and their colleagues (2009; 2007) conducted a series of experiments on the effect of IAC on the encoding and retention of visually presented information. In two of the experiments (Morgan et al., 2009, Experiment 1; Morgan et al., 2007, Experiment 1), the modified BWT task was adopted where participants were asked to replicate a series of colored blocks showing in the target window to the workspace window, and the IAC was manipulated following experiment 1 of Gray et al. (2006). Besides the usual measurement on interactive strategies, during the BWT task, a surprise memory test was conducted. In the memory task, the target and workspace windows were emptied, and participants were asked to recall the target pattern of the previous BWT problem. The results revealed a consistent shift from memory-less strategy to memory intensive strategy when IAC increased. Furthermore, the memory recall performance improved with an increasing IAC. The increasingly intensive memory-based strategy as a function of increasing IAC resulted in improved recall of the block pattern in the memory test. The authors concluded that the greater investment of effort to retrieve information presumably increased concentration on (and hence memory for) what was retrieved.

In two other experiments (Morgan et al., 2009, Experiment 2 & Experiment 3), the authors studied the effect of IAC on the memory recall after interruption. The task was similar to the one discussed above, except after 2, 3, 4, 5, or 6 correct copies of blocks, an interruption task involved solving a series of double-digit addition problem was inserted. The results showed that participants in the medium IAC condition were unable to resume the task without re-inspecting the target window. In contrast, high IAC participants were able to resume on a number of trials without inspecting the target window (Data in the low IAC condition was not available). Making information harder to access encouraged a memory-based strategy, which protected against some negative effects of interruption.

IAE/IAC has been represented through various surrogates in the above reviewed studies, including eye movement vs. mouse movement, varying levels of Fitts' index and varying levels of lookout time. Despite the differences in manipulation methods, a consistent finding was obtained- IAE/IAC is an important factor in determining information access behaviors. An increasing IAE/IAC encourages a more memory intensive strategy, and discourages access to KIW. We will describe the potential costs of this memory-based strategy below.

2.2.3 Metacognition and the prevalence of over-confidence bias

"When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it – this is knowledge."

--Confucius

The saying from the ancient Chinese philosopher demonstrates that knowledge is the ability of discriminating between knowing and not knowing. In fact 'knowledge' here refers to a specific area of research in today's scientific world - metacognition. Metacognition means the knowledge of one's knowledge and beliefs (Koriat, 2007). Under perfect calibration, metacognition reflects the exact level of one's knowledge and beliefs. However, over 40 years of empirical research suggests that people are subject to systemic biases and usually overconfident.

Overconfidence bias is ubiquitous in everyday life (Kahneman, 2011). For example, business start-ups were found to be unrealistically optimistic over the possibility of business success. They predicted an 81% probability of success for their start-up while the actual success rate was in the range of 30% to 70% (Koehler, Brenner, & Griffin, 2002). In addition, overconfidence bias has been reported in physicians' judgments of the likelihood of severe diseases (Wallsten, 1981), broadcasters' prediction on rare weather (Murphy & Daan, 1984), nonprofessional bridge players' judgments of their chances to win (Keren, 1987), consumers' judgments of products (Alba & Hutchinson, 2000), scientists' estimates of various physical constants (Henrion & Fischhoff, 2002), and pilots' judgments of their situation awareness (Sulistyawati, Wickens, & Chui, 2011).

Under the general umbrella of metacognition, there is an increasing body of research on the confidence judgment of information retrieved from memory. This is referred to as metamemory (Nelson & Narens, 1990). Two major domains in metamemory research are eyewitness memory (EM) and memory for general knowledge (GK), which will be discussed below.

It is a common assumption that a witness' subjective confidence is a good indication of his or her testimonial accuracy. The greater the confidence in memory recall and recognition, the more accurate it is. However, empirical research on EM contradict this common assumption; confidence is in fact a poor indicator of memory accuracy (Bornstein & Zickafoose, 1999; Krug & Weaver, 2005).

Bornstein & Zickafoose (1999) studied the relationship between confidence and accuracy in eyewitness reports using two experiments that imitated common situations in criminal cases. Participants were exposed to two confederates during the initial presentation of the scenarios. One week later they answered several questions concerning the characteristics of the two confederates such as the color of their clothes. The results revealed that participants were on average 19% overconfident. Similar results were reported in several other studies (Allwood, Granhag, & Johansson, 2003; Granhag, Stromwall, & Allwood, 2000; Olsson, 2000).

Besides the studies in criminal EM, Krug & Weaver (2005) studied the reliability of EM in civil cases. An experimental paradigm was proposed where participants were asked to follow recipes for a dish and later were given surprise memory tests on the specific brands of products used in the recipes. In addition, they were asked to report their subjective confidence for each response. The results showed that metamemory accuracy was poor after even a brief delay and subjects were overconfident in their responses. Furthermore, the overconfidence bias was exacerbated by longer delays.

Another domain where retrospective memory has been investigated is memory for general knowledge. Participants are usually presented with general-knowledge questions and asked to assess the probability that the chosen answer is correct. An overconfidence bias is typically observed, with mean probability judgments exceeding the proportion of correct answers (Lichtenstein & Fischhoff, 1982).

Lichtenstein and Fischhoff (1977, Experiment 3) studied confidence judgments on GK where participants responded to 75 questions with highly varied topics using a 4AFC (4 alternative forced choice) format and indicated their confidence to each answer. The results showed that the hit rates associated with confidence 0.50, 0.60, 0.70 and 0.80 were virtually identical, with all of them falling below the ideal calibration line. On average, participants were correct in 63.8% of the answers, but assigning a mean probability of 0.724, resulting in an 11.4% overconfidence. Along the same line, Bornstein & Zickafoose (1999) studied the relationship between confidence and accuracy in GK. The results showed that participants were on average 16% overconfident.

Overconfidence was also reported in research in classroom settings (Lundeberg, Fox, Brown, & Elbedour, 2000; Lundeberg, Fox, & Puncochar, 1994). Students were assessed on their learning results through course exams and asked to indicate their confidence on the correctness of their answers. An overconfidence bias was observed across several countries.

In general, a better confidence-accuracy relationship has been observed in GK than EM. A possible explanation was proposed that participants are able to use past experiences as benchmarks for assessing memory strength and weakness, whereas in EM, participants have little prior experience (Hollins & Perfect, 1997). Though very common, the findings of overconfidence bias are not universal. For example, Lichetenstein and Fischhoff (1977) found that overconfidence in GK questions tended to be stronger for more difficult questions. Indeed, when questions were extremely easy, underconfidence was observed rather than overconfidence.

In a word, a great amount of research in the area of retrospective memory has reported overconfidence bias in both EM and GK memory. The confidence judgment largely exceeds the objective accuracy.

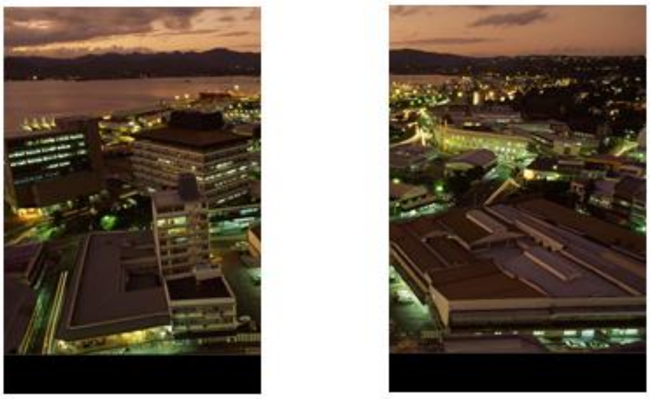
2.2.4 Factors affecting accuracy, confidence judgment and overconfidence

A variety of factors affecting accuracy, confidence judgment and overconfidence are summarized below, including **the accuracy-confidence inversion** in the study of Tulving (1981) (an important phenomenon to be deployed in study 2), hard-easy effect, expertise and practice, accountPresuccability, information accumulation, individual differences, and mood/depression. The review of the studies includes overconfidence bias in metamemory and decision making.

2.2.4.1 Accuracy-confidence inversion in Tulving (1981)

To study eye-witness memory, Tulving (1981) developed a basic experimental paradigm and found the pattern of accuracy-confidence inversion. In the experiment, participants studied a sequence of 160 pictures, denoted A, B, C…. Each picture was selected from the left or the right half of a bisected photo. The remaining other halves of the photos, denoted as A', B', C'…, were the similar counterparts of the studied pictures, and served as distracters in later recognition tests (Figure 2-2). After viewing the 160 pictures, the participants performed a two-alternative forced-choice (2AFC) test with different conditions. In condition AA', the target picture (A) was paired with its similar counterpart A' (Figure 2-3). 2. In condition BC', the target picture (B) was paired with a picture (C') which was similar to another studied picture (Figure 2-4).

Figure 2-2 Illustration of picture A, B, C and their similar counterparts A', B' and C'.



A A'

Figure 2-3 Illustration of picture type AA'

B C'

Figure 2-4 Illustration of picture type BC'

Following each selection, the participants rated their confidence of the answers on a 1 to 4 scale. The results revealed the pattern of accuracy-confidence inversion, as illustrated in Table 2-3.

Table 2-3 Hit rate, false positive rate and confidence (Source: Tulving, 1981)

Picture type	**Targets**		**Distractors**	
	Hit rate	Confidence	False positive rate	Confidence
AA'	0.71	2.05	0.29	1.55
BC'	0.58	2.40	0.42	2.28

The participants discriminated targets from distractors better in condition AA' than condition BC'. The recognition accuracy was 0.71 in condition AA' versus 0.58 in condition BC'. Despite higher accuracy, however, participants exhibited lower confidence in condition AA' than condition BC', in both hit's confidence (2.05 vs. 2.40) and false positives confidence (1.55 vs. 2.28) (Significance value were not available). Thus there was a negative correlation between recognition accuracy and confidence between the two conditions. Although the measures of accuracy and confidence did not use the same scale as in Tulving (1981), the results imply that participants were more overconfident in condition BC' than condition AA'. The same accuracy-confidence inversion pattern was replicated in later studies (Dobbins, Kroll, & Liu, 1998).

2.2.4.2 Hard-easy effect on general knowledge questions

The hard-easy effect was first reported in a study of Lichtenstein & Fischhoff (1977) and afterwards replicated in many studies (Lichtenstein & Fischhoff, 1980; Pulford & Colman, 1997). In a brief, the hard-easy effect shows that as the difficulty of general knowledge questions increases, so does overconfidence.

Difficulty levels in these studies can be post-defined or pre-defined. In a study with post-defined difficulty, difficulty was measured by the subjects' performance after they answer a set of questions. If the percentage of correct responses for a question was below a threshold, it is grouped as a hard question; otherwise an easy question. By doing this, the set of questions was divided post hoc into hard vs. easy groups. Experiment 3 of the seminal paper (Lichtenstein & Fischhoff, 1977) adopted the post-defined method. In the experiment, subjects answered 70 general knowledge questions with highly varied topics (e.g. Oslo is the capital of (a) Norway, (b) Finland). Questions were sorted into two groups according to the percentage of subjects answering them correctly: easy items (67% or more subjects were correct) and hard items (less than 67% subjects were correct). The results (Table 2-4) reported that subjects were less accurate but more overconfident in hard questions.

Table 2-4 Summary of studies on hard-easy effect (OC – overconfidence)

Study/Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
Lichtenstein & Fischhoff, 1997, Exp 3 (GK)	Easy	80	N.A	76	N.A	-2.3	N.A
	Hard	50		69		19.3	
Lichtenstein & Fischhoff, 1980 (GK)	Easy	Detailed numbers not available.					
	Hard						
Lichtenstein & Fischhoff, 1997, Exp 5 (GK)	Easy	80	N.A	78	N.A	-2	N.A
	Hard	62		74		12	
Pulford & Colman, 1997 (GK)	Easy	79.0	P< .001	67.6	P< .001	-11.4	P< .001
	Medium	57.5		48.0		-9.5	
	Hard	21.0		28.5		7.5	

In other studies, difficulty was manipulated as a priori. Lichtenstein and Fischhoff (1980) developed a set of 500 general questions whose difficulty could be defined independently. Two example questions could be, Is Melbourne further from Rome or from Tokyo? Or Is Melbourne further from Rome or from Athens? Since the distances from Melbourne to Rome, to Tokyo and to Athens can be measured objectively, the ratio between the two distance numbers in a question was calculated as the difficulty of the question. The study reported overconfidence in hard questions and underconfidence in easy questions. Experiment 5 of Lichtenstein & Fischhoff (1977) and a study of Pulford and Colman (1997) adopted similar methods to manipulate difficulty. Their results consistently showed the hard-easy effect.

In a word, there are consistent findings of the hard-easy effect on accuracy and overconfidence. As questions become harder, accuracy decreases and overconfidence increases.

2.2.4.3 Expertise and training

Effect of expertise is closely related to the hard-easy effect, for a fairly obvious reason that when subjects gain more experience in a task, they usually find the task become easier, and therefore extrapolating from the hard-easy effect, overconfidence should diminish.

Experiment 2 of Lichtenstein and Fischhoff (1977) asked subjects to decide whether a handwritten message was made from European or American adults. Half of the subjects received training on this task. In the training phase, they studied 10 handwritten messages with feedback. The other half received no training. The results (Table 2-5) showed higher accuracy, higher confidence rate but lower overconfidence (Significant level not reported) in subjects with training.

Table 2-5 Summary of studies on effect of expertise (OC – overconfidence)

Study/Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
Lichtenstein & Fischhoff, 1997, Exp 2 (DM)	Trained	71.4	N.A.	77.9	N.A.	8.5	N.A.
	Untrained	51.2		65.3		14.1	
Keren, 1987 (DM)	Expert	56	N.A.	51.4	N.A.	-4.6	P< .05
	Novice	60		70.7		10.7	
Lichtenstein & Fischhoff, 1997, Exp 8 (GK)	Psychological knowledge (expert)	75.7	No diff.	78.0	No dif.	N. A.	N. A
	General knowledge (novice)	76.0		77.8		N. A.	

Keren (1987) studied the differences in calibration between expert and novice bridge players. In two experiments, expert and amateur bridge players were asked to estimate the probability that the final contract would be made. The calibration curve (Figure 2-5) showed that expert players were almost perfectly calibrated whereas amateurs were overconfident. Also, novice players were significantly more overconfident than expert players.

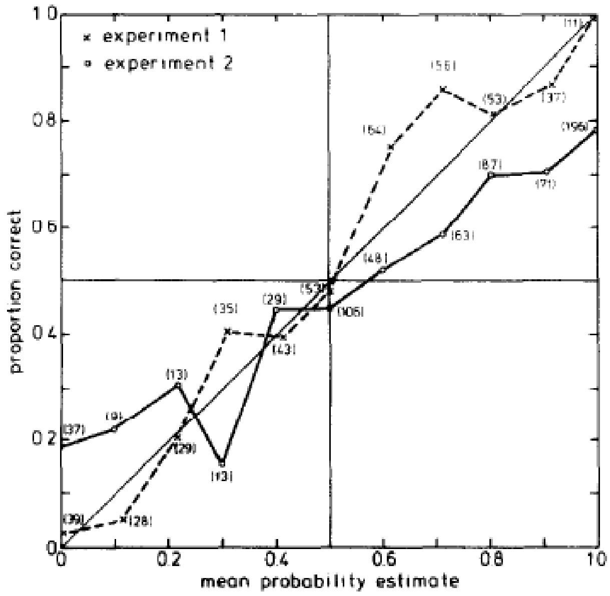


Figure 2-5 Calibration curves for expert (Experiment 1) and amateur (Experiment 2) players (Source: Keren, 1997)

Despite the general support for the effect of expertise, some studies showed no difference between expert and novice in confidence judgments. Experiment 8 of Lichtenstein and Fischhoff compared the performance of the same group of subjects on general questions and psychological questions. Since the subjects were from department of psychology, it was expected that the subjects would have higher performance but lower overconfidence with psychological questions. However, the study found no difference in accuracy, confidence, and overconfidence. Furthermore, it is certainly the case that experts in a wide range of professions such as stock forecasters and political scientists, have continued to show substantial levels of overconfidence (Kahneman, 2011).

In sum, the effect of expertise and training on overconfidence has been consistently reported. When subjects become more experienced, they become less overconfident. However, in the results on accuracy and confidence judgments show opposite directions - In Experiment 2 of Lichtenstein and Fischhoff (1977), more experienced subjects showed higher accuracy and confidence than less experienced users, whereas in the study of Keren (1987), more experienced subjects showed lower accuracy and confidence than less experienced ones.

2.2.4.4 Effect of accountability on decision making and general knowledge

Accountability was defined as the "social pressure to justify one's views to others" (Tetlock & Kim, 1987). It has been posited that people under certain conditions, can be motivated to carry out more complex, self-critical and effort-demanding information processing (Borgida & Howard-Pitney, 1983). Thus, it was expected that accountable decision makers have higher possibility of adopting cognitive complex rules and thus reduce decision making biases such as overconfidence (Tetlock & Kim, 1987).

Tetlock & Kim (1987) studied the effect of accountability in a personality prediction task. In the task, subjects were presented with an individual's responses to 16 questions in a personality test, and asked to write a personality sketch of that individual. After that, subjects were asked to predict the individual's answers to another set of 16 questions on personality and indicate their confidence. Accountability is manipulated in three conditions: no-accountability – subjects were assured before the experiment that their predictions were completely confidential and untraceable; pre-exposure accountability – before studying the PRF test, subjected were told that after the experiment they would later participate in an interview to justify their answers; post-exposure accountability – subjects were told the same instruction as the pre-exposure condition, only after they read the PRF test.

The analysis results showed higher hit rate and smaller overconfidence in accountability conditions. The confidence levels were not reported in the study, it seems, however, that there was no significant difference between confidence judgments.

Although not explicitly adopting the term “accountability”, another two studies investigated how overconfidence changes when a decision has to be made. Sniezek, Paese, and Switzer (1990) studied the effect of whether subjects had to make a decision explicitly on confidence judgment of responses to general knowledge questions. There were three experimental conditions: choice-positive – subjects chose an answer of a GK question and indicate confidence; arbitrary cue-positive – the answer to a question was arbitrarily chosen and subjects were asked to indicate the confidence on the chosen answer; uncued-positive – subjects were asked to indicate the confidence without choosing a correct answer (To calculate hit rate in this condition, following the completion of the probability assessment, subjects were asked to choose the correct answer of the same questionnaire). The results showed that high accountability condition (choice positive) had higher hit rate, lower confidence and less overconfidence than the other two conditions.

Table 2-6 Summary of studies on effect of accountability (OC – overconfidence)

Study /Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
Tetlock & Kim, 1987 (DM)	No accountability	59.1	P< .01	72.7	N.A.	13.6	P< .01
	Pre-exposure accountability	69.8		73.8		4.0	
	Post-exposure accountability	62.3		71.9		9.6	
Sniezek, Paese, & Switzer, 1990 (GK)	Uncued	55	P< .05	70	P<0.05	15	P< .001
	Arbitrary cue	52.5		73.5		21	
	Choice	57.5		67		9.5	
Ronis & Yates,	Arbitrary cue	64.6	No diff.	81.2	P <0.001	16.5	P< .001

1987 (GK)	Choice	66.4		76.2		9.8	
Ronis & Yates, 1987 (DM)	Arbitrary cue	58.9		67.6		8.7	
	Choice	60.1		66.7		6.6	

Ronis & Yates (1987) used the same research paradigm to study the effect of making choice on general knowledge and prediction of basketball winning team. The results showed subjects were more confident and more overconfident in the arbitrary cue condition. Although detailed statistics were not reported for GK and DM alone, the data suggests that the effect was stronger in the GK condition.

In summary, the results showed a consistent trend that as accountability increases, accuracy increases and overconfidence decreases.

2.2.4.5 Effect of information on decision making

Does accumulating more information help decision making? This question was addressed by a few studies, with mixed results. Oskamp (1965) investigated how psychologist's confidence judgments in their clinical decisions changed as they accumulated more patient information. In the study, the patient information was revealed gradually in four stages. After each stage, subjects responded to 25 questions of the patient and reported their confidence levels. The results revealed an increase in confidence (and very likely overconfidence) without change in accuracy.

Table 2-7 Summary of studies on the effect of information

Study /Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
Oskamp, 1965 (DM)	Stage 1	26	No diff.	33.2	P < .001	7.2	N.A.
	Stage 2	23		39.2		16.2	
	Stage 3	28.4		46.0		17.6	
	Stage 4	27.8		52.8		25.0	
Peterson & Pitz, 1988 (DM)	One cue	41	P<0.05	48	P< .01	7	N.A.
	Two cues	47		55		8	
	Three cues	53		64		11	

However, another study showed that an increase in number of cues used to predict the number of game won by a National League baseball team resulted in both increasing accuracy and confidence (Peterson & Pitz, 1988). Although the above mentioned two studies did not report the changes of overconfidence as subjects accumulated more information, the data suggested a consistent trend that as the amount of information increases, overconfidence increases.

2.2.4.6 Individual difference on memory performance

Individual differences in confidence judgments have received quite an amount of attention, yet the results are not conclusive. Some studies suggest that miscalibration (or bias) may have a trait-like quality whereas others offer opposite evidence. Studies from each side will be discussed below.

West & Stanovich (1997) investigated subjects' accuracy and confidence in two different domains: general knowledge and motor performance. The test on general knowledge followed the well-established paradigm where subjects answered questions in a two-choice format and indicated their confidence level. The test of motor skills involved a penny slide task where subjects attempted slide pennies onto a strip on the table. The subjects performed 2 trials of the penny slide task and prior to each trial predicted their performance. The authors divided the subjects into two groups on the basis of a median split of their overconfidence scores of their general knowledge tests. The results showed that subjects with high overconfidence scores in general knowledge also demonstrated high overconfidence scores in the motor task. There was a significant correlation ($P<0.01$) between the magnitude of overconfidence between the two tasks. Along the same line, Bornstein & Zickafoose (1999) studied the accuracy-confidence relationship in general knowledge and eyewitness memory. The author calculated the correlation across domains for each participant and found a significant positive correlation in confidence judgment and overconfidence.

Table 2-8 Summary of studies on individual differences (OC – overconfidence)

Study/Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
West & Stanovich, 1997, (GK)	High OC	61.8	P< .001	79.2	P< .001	17.2	P< .001
	Low OC	67.9		71.1		5.3	
West & Stanovich, 1997, (Motor task)	High OC	108.8	No diff.	122.2	P< .05	13.4	P< .05
A different scales were used	Low OC	108.3		107.7		-0.6	
Bornstein & Zickafoose, 1999 (GK & EM)	GK	55	Not sig. corr	74	Corr= .33, p< .01	19	Corr = .34, p< .01
	EM	31		47		16	
Hosch & Platz, 1984 (EM)	Not applicable. The study used correlation statistics.						
Blais, Thompson, & Baranski, 2005 (Vocabulary, GK ,perceptual)	Not applicable. The study treated cognitive styles as covariates in the analysis.						

Evidence against overconfidence as a trait-like quality includes the failures to associate a variety of personal traits with overconfidence. Hosch & Platz (1984) studied the association between a subject's self-monitoring score ("a person's concern about behaving in an appropriate way in social situations") and eyewitness memory performance. The results revealed a significant association between SM and accuracy ($r=0.51$, $p<0.001$). The higher the level of SM, the more likely the subject was to be accurate in his or her identification. However, there was not a significant association between SM and confidence judgment ($r=0.11$, n.s.)

Blais, Thompson and Baranski (2005) explored the association between cognitive style ("a person's chronic motivations that principally determine that initiation, course, and cessation of information seeking and processing") and confidence judgments in three tasks: the vocabulary task required participants to choose which of two potions was a closer match to a target word; the general knowledge task required participants to select one of two options to a question; the perceptual comparison task asked participants to determine which of two horizontal lines was longer. Four cognitive style measures were used: Need for cognition scale (NFC), Desire for structure (DFS) and Response to a lack of structure (RLS), and Personal fear of invalidity scale (PFI). The results revealed a significant main effect of NFC on accuracy, with higher NFC score predicting a higher average accuracy. However, there was no significant effect of any of the cognitive styles measures on the confidence and overconfidence indices.

In general, the above results are hard to interpret. The studies aiming to link personality traits to confidence judgment are generally not successful.

2.2.4.7 Mood/depression

Dunning and Story (1991) conducted two experiments to study the difference in predicting future events between depressed and nondepressed individuals. In both experiments, subjects were asked to predict whether they would experience a series of events during the course of a semester. After making a prediction, they were also asked to estimate the probability that each prediction would prove true. Afterwards, subjects responded to a test which measured their depression scores. If their scores were above certain threshold, they were classified as depressed otherwise nondepressed. The experimental results showed that depressed individuals had low accuracy but higher overconfidence.

Table 2-9 Summary of studies on effect of mood

Study/Experiment	IV	Hit rate	Sig.	Conf.	Sig.	OC	Sig.
Dunning and Story, 1991, exp 1 (DM)	Depressed	72.4	P< .005	83.8	No diff.	11.4	P< .005
	Nondepressed	77.5		83.5		6.1	
Dunning and Story, 1991, exp2 (DM)	Depressed	73.9	P< .05	82.7	No diff.	8.8	No diff.
	Nondepressed	76.7		83.5		6.3	

2.2.5 Tradeoffs between accuracy and effort

Section 2.2.1 and 2.2.2 reviewed the concept of effort and specifically, the information access cost and found that during decision making people are sensitive to low level costs. Even a one second change in temporal cost alters one's interactive strategy. Also, Section 2.2.3 described the research on metacognition and found that a person's knowledge of his knowledge is subject to systemic distortions and such distortions can alter his follow-up actions. For example, an overconfident diagnosis may prevent a physician from considering other possible diagnoses and thus adopt a non-optimal treatment plan.

Indeed, a number of studies from decision making research have proposed models and hypotheses that integrate the concepts of effort and knowledge (or metacognition) to predict human decision behavior. These models posit a cost-benefit theory, that in selecting among decision strategies, the decision maker balances a strategy's accuracy (or perhaps his perception of accuracy) against its demand for effort. The strategy that maximizes the benefit over cost is selected (Payne, Bettman, & Johansson, 1993)

Wickens et al. (2004) compared the relationships between effort and performance for two types of decision making (DM) strategies using the performance-resource function as shown in Figure 2-6. 'Algorithm' refers to a DM strategy leading to optimal performance, granted maximal investment of effort, whereas 'heuristic' refers to a DM strategy resulting in good enough performance with lesser effort. When choosing a decision strategy, people evaluate the utility of that strategy – the weighted sum of effort and performance.

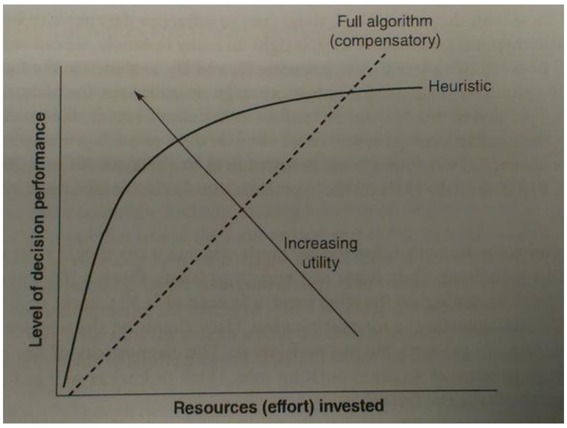


Figure 2-6 Performance resource function (Source: Wickens et al., 2004)

Along the same line, Payne, Bettman and Johnson (1988) proposed a contingent decision model which illustrates the relationship between effort and accuracy among DM strategies. In the model, the authors identified six 'elementary information processing' (EIP) mechanisms in assessing the effort investment of a decision strategy: Read, Add, Compare, Multiply, Difference, and Eliminate. The model characterizes a variety of DM strategies into compensatory and non-compensatory strategies. Compensatory strategies require the decision maker to trade off attribute values, thus requiring more EIPs and leading to greater accuracy. Non-compensatory strategies, on the contrary, generally do not make use of all available information, and therefore require less EIPs and lead to lower accuracy. The model predicts that the selection of a particular choice is determined by the trade-off between desired accuracy (more EIPs required) and available effort (less

EIPs afforded). Two follow-up experiments confirmed the contingent decision model that people were highly adaptive in their decision behaviors. When time pressure increased (available effort decreased) subjects switched from compensatory strategies to more non-compensatory strategies.

In early research on the cost-benefit theory of decision strategy, researchers measured deductively the accuracy and the effort for a DM strategy. However, it was argued that decision makers' perceptions of accuracy and effort, instead of the objectively researchers' deductively derived accuracy and effort, are the determinants of strategy selection (Chu & Spires, 2003; Fennema & Kleinmuntz, 1995; Payne et al., 1993). Chu and Spires (2003) conducted a survey to study perception of decision strategies and showed that participants' perception of the accuracy and effort of various decision strategies largely agree with researchers' deductions. Bettman, Johnson and Payne (1990) studied post-decisional perceptions of effort associated with various decision strategies. They reported that participants' perceptions of experienced effort in experiments were highly consistent with objectively deduced measures of effort based on EIPs. Similarly, Fennema and Kleinmuntz (1995) found that the decision makers' perception of experienced effort and experienced accuracy were well calibrated according to task difficulty. However, Fennema and Kleinmuntz found that decision makers had only limited ability to anticipate effort accuracy attained for given tasks. The poor calibrated anticipated accuracy and anticipated effort impose challenges for decision makers who use these values that may be spurious or biased to determine strategy selection.

In summary, section 2.2 reviews theories and studies pertaining to the information retrieval behavior during pre-handover. There is strong empirical evidence for the effect of information access effort on information retrieval strategies. However, the effect of perceived accuracy, anticipated to be another determinant of information retrieval strategies, has not been rigorously studied. To fill the research gap, chapter 3 presents a research model integrating the two factors, arguing that the information retrieval behavior and the performance are determined by the information access cost and the confidence judgment (perceived accuracy). Moreover, the optimal strategy is obtained by the cost-effective balance between information access cost and confidence judgment.

2.3 Handover media in communication

Section 2.3 reviews two classes of literature relevant to the study of handover communication media. It begins with a review of the common ground theory and mediated communication. Common ground theory was proposed from the social perspective of communication and applied widely in the study of mediated communication. A handful of recent studies, however, suggested its limitation to account for some research results. Therefore, the second class of literature introduces the theory of individual-task-technology fit, with a focus on the "fit" between the communication task and the communication technology, arguing that in order for a technology to benefit communication processes and outcomes, the "fit" between the technology and the communication task should be taken into consideration.

2.3.1 Common ground theory and technology mediated communication

2.3.1.1 Common ground theory and media affordances

To study the effect of communication media on handover communication, an appropriate framework is needed to model the relationships between communication media and communication outcomes. A widely applied framework is the mathematical theory of communication proposed by Shannon and Weaver (1949). In their view, communication comprises an information source that produces a message, a transmitter that operates on the message to create a signal, a channel through which the message is transmitted, a receiver which transforms the signal back and a destination which/who receives the message. This communication model is widely used in information communication industry and has produced many useful applications. For example, the use of different frequencies to transmit signals draws from the “channel” concept and makes it possible for us to listen to different radio channels and watch various TV programs.

However, communication is not simply a matter of exchanging linguistic codes. Rather, people communicate to influence what others know, believe, or do (Austin, 1960). It is a social processes for producing and understanding messages, and constructing a shared understanding (Morrow & Fisher, unknown). A collaborative model of communication was then proposed by Stanford psychologist Herbert Clark (1996). He stated that communication participants need to update their understanding of a subject moment by moment and finally achieve a shared understanding of the information they intend to communicate (Clark & Brennan, 1991). When A presents utterance u to B, it will result in one of four possible states.

State 0: B did not notice that A uttered any u.
State 1: B noticed that A uttered some u.
State 2: B correctly heard u.
State 3: B understood what A meant by u.

It has to be noted that there exists a remarkable resemblance between the common ground theory and the concept of team situation awareness (TSA). TSA was defined as team members’ understanding of the system/task status and of the team status (Endsley & Jones, 2001; Salas, Prince, Baker, & Shrestha, 1995) which is achieved by cognitive processes, such as perception, comprehension and projection, and additional activities such as communication and coordination (Endsley, 1995). In the context of clinical handovers, common ground and TSA of a patient is achieved mainly by the information exchanged between the handover senders and receivers.

In common ground theory, Clark proposed that the process of establishing common ground changes with the communication medium and each medium can be characterized by its affordances (Brennan & Lockridge, 2006). For example, video teleconference allows communication parties to see each other's spontaneous facial expressions, whereas normal telephone call does not. Past studies have proposed four major affordance sets (Table 2-10), which are reviewed below individually (Clark & Brennan, 1991; Daft & Lengel, 1984; Daft, Lengel, & Trevino, 1987; Dennis & Valacich, 1999; Kraut, Fussell, Brennan, & Siegel, 2002; Whittaker, 2002).

Table 2-10 Summary of four affordance sets (Categories and subcategories are grouped according to the definition)

Common ground theory (Brennan & Lockridge, 2006; Clark & Brennan, 1991)	**Theory of media richness (Daft & Lengel, 1984; Daft et al., 1987)**	**Theory of media synchronicity (Dennis & Valacich, 1999)**	**Whittaker's affordance classification (Whittaker, 2002)**
Sequentiality: participants take turns in an orderly fashion	*Feedback*: the ability of allowing instant feedback	*Immediacy of feedback*: the extent to which a medium enables users to give rapid feedback	*Interactivity*: the ability to feedback via backchannels, completions, interruptions
Cotemporality: messages are received without delay			
Simultaneity: participants can send and receive at the same time			
Co-presence: participants share the same physical environment, including a view of what each other is doing or looking at	*Multiple cues*: the ability of supporting multiple cues	*Symbol variety*: the number of ways in which information can be communicated. It combines Daft and Lengel's multiplicity of cues and language variety.	*Modality*
Visibility: participants can see each other, but not necessarily what the other is doing or looking at	*Language variety*: the ability of supporting range of meaning that can be conveyed with language symbols		
Audibility: participants communication by speaking			
	Personal focus: the ability of conveying personal feelings and emotions		*Expressivity*: technologies' ability of describing global and affective information.
		Parallelism: the number of simultaneous conversations that can exist effectively	

Revisability: messages can be revised before being sent		*Rehearsability*: the extent to which a medium enables the sender to rehearse or fine tune the message before sending.	
Reviewability: messages do not fade over time		*Reprocessability*: the extent to which a message can be reexamined or processed again within the context of the communication event.	*Context*: technologies' ability of supporting conversational context

Clark and Brennan (1991) outlined eight types of affordances that various media provide: copresence, visibility, audibility, cotemporability, simultaneity, sequentiality, reviewability, and revisability (Table 2-11).

Table 2-11 Six media and their associated affordances

	Copresence	Visibility	Audibility	Cotemporality	Simultaneity	Sequentiality	Reviewability	Revisability
Face to face	○	○	○	○	○	○		
Telephone			○	○	○	○		
Video conference		○	○	○	○			
Answering machine			○				○	
Email							○	○
Letters							○	○

Copresence: participants share the same physical environment, including a view of what other participants are doing or looking at. In co-located group work, participants are usually in the same environment and able to view what other participants are doing. This is not available in distributed group work.

Visibility: participants can see each other, but not necessarily what the other is doing or looking at. In face to face communication or video conferencing, participants are able to observe the eye gaze, facial expression and gestures of the other parties.

Audibility: one participant can hear the others. Face to face, video conference, telephone offer audibility, while email and keyboard conferencing do not.

Cotemporality: messages are received without delay. In verbal communications, once an utterance is produced, it is heard by the addressee immediately. This is not the case in emails and voice mails.

Simultaneity: participants can send and receive at the same time. In face to face communication, one can nod and make facial expression while listening to the speech of others. In conversation using walkie-talkie, one can only either speak or listen at certain time.

Sequentiality: participants take turns in an orderly fashion. One turn's relevance to another is signaled by adjacency. In face to face conversation, sequential turns are naturally formed. With email and answering machines, a message may be separated by any number of irrelevant messages.

Reviewability: messages do not fade over time. In media such as email and audiotape, messages can be viewed and heard after being communicated, whereas in verbal communication, messages fade quickly.

Revisability: messages can be revised before they are sent. In everyday verbal communications, participants do not revise what they intend to say before sending out the message, while in email and keyboard conferencing they can check the sentences before pressing the send key.

Kraut, Fussell et al. (2002) studied the effects of proximity on collaboration. They adapted Clark and Brennan's framework and expanded the list to include two additional affordances: tangibility (Participants can touch other people and objects in the environment) and mobility (People can move around in a shared environment).

Though not using the terminology “affordance” explicitly, research in organizational communication was among the earliest to study the dimensions that characterize a communication medium. Theory of ‘media richness’ and theory of ‘media synchronicity’ were developed to study how organizations utilize various communication tools in daily operations. Theory of media richness was proposed by Daft and his colleagues (Daft & Lengel, 1984; Daft et al., 1987). Richness is defined as the capability a medium to facilitate insight and rapid understanding, which is determined by four factors: feedback, multiple cues, language variety, and personal focus. Feedback is a medium’s ability of allowing instant feedback. Telephone calls allow the communication participants to provide instant feedback to each other while emails do not. Multiple cues is the ability of supporting multiple cues such as body gestures and voice inflection. Language variety means the range of meaning that can be conveyed with language symbols. Numbers convey greater precision of meaning than natural language. Personal focus is the ability of conveying personal feelings and emotions.

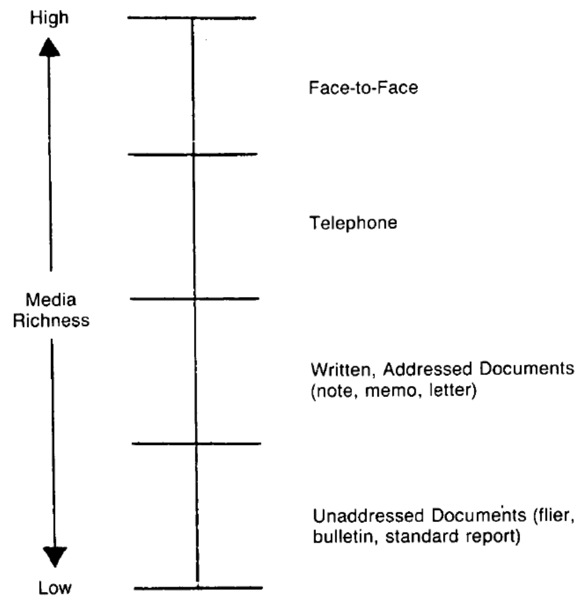


Figure 2-7 Hierarchy of media richness (Source: Daft & Lengel, 1984)

According to theory of media richness, each medium has its unique position along the richness continuum (Figure 2-7). Face to face communication is considered the richest communication medium. It allows rapid mutual feedback, simultaneous communication of multiple cues. Also, it uses a high variety of natural language and conveys personal feelings and emotions. On the other end of the continuum, formal, unaddressed documents are the lowest in media richness. These documents often use numbers to communicate quantifiable information but do not have the information carrying capacity of natural language. In addition, they do not allow multiple cues nor focus towards any individual.

Dennis and Valacich (1999) proposed the theory of media synchronicity. Synchronicity is defined as a "state in which individuals are working together at the same time with a common focus". The ability of media to support media synchronicity is influenced by five intrinsic capabilities of the media: immediacy of feedback, symbol variety, parallelism, rehearsability, and reprocessability. Immediacy of feedback is the extent to which a medium enables users to give rapid feedback. Symbol variety is the number of ways in which information can be communicated. It combines Daft and Lengel's (Daft & Lengel, 1984) multiplicity of cues and language variety. Parallelism refers to the number of simultaneous conversations that can exist effectively. With a telephone, only one conversation can take place at one time. In contrast, many electronic media enable many simultaneous conversations. Rehearsability stands for the extent to which a medium enables the sender to rehearse or fine tune the message before sending. Reprocessability is the extent to which a message can be reexamined or processed again by the receiver within the context of the communication event.

The last affordance set was proposed by Whittaker (2002) by summarizing the theories and methods in technology mediated communication. He stated that although complex set of affordances were proposed, the majority of existing studies can be explained by two affordances: modality and interactivity.

Table **2-12** shows Whittaker's taxonomy of communication media and their associated affordances. The author also emphasized the necessity to enrich the set of affordances by adding two affordances: context – technologies' ability of supporting conversational context, and expressivity – technologies' ability of describing global and affective information.

Table 2-12 Interactivity and Modality (Source: Wittaker, 2002)

Affordance		**Interactivity**	
		Interactive	Non-interactive
Mode	Linguistic	Phone, Audio conference, Chat instant message	Email, Answering machine, Voicemail, Fax, Letter, Usenet
	Linguistic and Visual	Videoconference, Videophone, Shared workspace	Video mail

2.3.1.2 Technology mediated communication

Mediated communication is ubiquitous today. People rely on it to extend the limits of human communication, which enables communication at distance and across time. The fundamental goal of studying technology mediated communication is "to explain the relationship between the affordances of different mediated technologies and the communication that results from using those technologies"(Whittaker, 2002). The following literature presents studies examining the effects of six affordances, including reviewability, audibility, visibility, shared workspace, interactivity and revisability.

It has to be noted that communication tasks differ on the degree to which all participants are able to incrementally contribute to and modify the content (Oviatt & Cohen, 1991). Toward one extreme, participants can contribute fairly equally to a communication task, such as in a negotiation task. Towards the other extreme, a communication task could be extremely uni-directional. For instance, in a military commander-gunner task, the commander makes the majority of the communication by directing the gunner and the latter mainly acknowledges the reception of information. Past studies on clinical handover showed that handover senders play a more active role in the communication by transferring critical patient information to handover receivers. However, the communication was by no means uni-directional, because handover receivers were found to constantly clarify doubts and prompt for more information (Ye et al., 2007). Therefore, this section covers communication studies from both ends, with a focus on research more towards the uni-directional end.

Reviewability was defined as the extent to which a message can be reexamined or processed again within the context of the communication event and audibility was defined as whether participants communicate by speaking (Clark & Brennan, 1991; Dennis & Valacich, 1999). Below the two affordances are examined and discussed together for three reasons. First, reviewability and audibility are closely related to each other - auditory messages are usually more “transient” and hence affords lower reviewability than text-based messages, and vice versa. Second, existing research in technology mediated communication did not clearly disentangle the two affordances. Usually, participants were asked to communicate by either speech without recording (low reviewability) or text-based communication with the ability of referring back to previous messages (high reviewability). Third, there is a growing body of research examining the redundancy effect of presenting the same information in both an auditory channel and a text-based channel (Lu et al., 2013).

A body of research on the effect of audibility and reviewability originated from the dual-task performance paradigm, where participants engaging in an ongoing task are intersected by an interrupting task through visual, auditory, tactile interfaces. Helleberg and Wickens (2003) compared flight pilots’ performance of following Air Traffic Control (ATC) instructions using a text-based display, a synthesized voice and a redundant voice-text format. In the task, pilots flew simulated cross-country flights for the ATC clearance information. After receiving each ATC instruction (e.g. “Climb to flight level 230”), pilots read back while aviating (flight path control) and navigating (spotting traffic in sight). Communication quality was assessed by the proportion of ATC readback errors and the results showed a significantly smaller number of errors using the text-based display and redundant voice-text display, compared to the voice display. The visual advantage could have resulted from the affordance of reviewability – the operator can use vision to quickly scan the forgotten information especially when the length of message is challenging working memory (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003; Wickens & Gosney, 2003). In two recent studies, Wickens and colleagues (Lu et al., 2013; Wickens, Prinet, Hutchins, Sarter, & Sebok, 2011) conducted meta-analyses of 68 and 29 studies on task performance as a function of interrupting task modality. The two studies showed that the communication accuracy of the interrupting tasks (which is of interest for the dissertation) enjoyed a significant redundancy gain against the best single modality component (almost always auditory). Moreover, comparing two types of auditory displays showed that a relatively permanent (with a repeated tone) auditory display was better than a highly transient auditory display; hence amplifying the importance of reviewability.

Besides the studies from the dual-task performance paradigm, research from small group communication examined the effect of audibility on collaborative communication tasks. Newlands, Anderson, and Mullin (Newlands, Anderson, & Mullin, 2003) studied the performance of using text-based communication and speech over a series of collaborative map tasks. The task was completed by two participants, one instruction giver and one instruction follower. The aim of the task was for the instruction giver to tell the instruction follower about a route, so that the instruction follower could reproduce it as accurately as possible. Task performance was accessed by the accuracy of the route reproduced and the time to complete a task. The results showed that accuracy of text-based communication was initially poorer compared to speech, but improved as the participants gained experience. Obviously, it took participants in the text-based communication far longer to complete the task than participants who could speak to each other. Along the same line, Chapanis, Ochsman, Parrish, and Weeks (1972, 1977) compared the communication efficiency of two cognitive tasks using four types of communication modes: face-to-face communication, audio only, remote handwriting, and remote typewriting (Details about the two cognitive task is explained later) and found that problems took longer to solve in the two written modes.

In a sum, the research on reviewability/audibility provides evidence for the effect of reviewability on communication accuracy and the effect of audibility on communication efficiency.

Visibility was defined as the extent to which communication participants can see each other, but not necessarily what the other is doing or looking at (Clark & Brennan, 1991). There is a considerable amount of research studying the effect of visibility on communication processes and outcomes.

Positive advantages of visibility have been found in studies focusing on communication technologies' ability of supporting the transmission of feelings, emotions, and attitudes. Stephenson, Ayling and Rutter (1976) compared the impact of speech only and face to face communication in a debating task. The results showed that communication using speech only was more impersonal and task-oriented. In addition, participants believed that technologies providing visual information through video/speech or face to face were better than telephone only since it is important to access interpersonal information in a debating task. Similarly, Rutter (1987) showed that compared to face to face discussion, speech discussions were perceived to be more socially distant. The positive advantage of visibility was reported in other studies involving negotiation (Short, 1972), achieving consensus (Hiltz, Johnson, & Turoffm, 1986) and coalitions (Reid, 1977).

The positive evidence of visibility was consistent with the bandwidth hypothesis proposed during early works on technology medicated communication. The bandwidth hypothesis focused on the differences between modes supported by various communication technologies and those by face to face communication and posited that the closer a mode resembled face-to-face communication, the more effective and efficient the communication was (Williams, 1977). However, the advantage of visibility was disapproved by some studies comparing different communication modes. Chapanis et al. (1972, 1977) compared four types of communication modes: face-to-face communication, audio only, remote handwriting, and remote typewriting on two cognitive tasks. On the first task, participants were assigned to jointly build a mechanical object. One participant was given the instructions and the other the unconstructed mechanical components. On the second task, participants were asked to find a location on a map satisfying a list of preset criteria. Similarly, one participant was given the list of criteria and the other the map. Communication efficiency was measured by time to find solutions. The results showed that problems took longer to solve in the two written modes. However, there was no significant difference between audio only and face to face mode. Thus, adding visual information to speech did not necessarily lead to more efficient communication.

In another study, Clark and Krych (2004) studied efficiency and accuracy when a director and a builder worked together to assemble 10 simple Lego models. The efficiency was assessed by the time used to finish a Lego model and accuracy was assessed by the number of wrongly placed Lego blocks. A block was counted as in error whenever it was the wrong color or size, or in the wrong location or orientation. The director had a prototype for each model and gave instructions and the builder followed the instructions and did the physical work. The 10 Lego tasks were performed using five methods, 4 interactive and 1 non-interactive (the performance difference between interactive and non-interactive conditions will appear in the discussion of interactivity). The 4 interactive methods were further divided into four conditions in a two factor design, defined by whether the faces and the workplace each were visible or not. In half of the conditions, the two partners could see each other's faces, and half the time they could not. The results showed no significant difference in building time and assembly error when the faces were visible.

In sum, for more emotion oriented tasks, technologies that transmit visual information such as facial expressions and gaze may be critical; for more cognition oriented tasks, providing participants with such visual information did not necessarily lead to higher effectiveness and efficiency.

Co-presence/shared workspace was defined as the extent to which participants share the same physical environment, including a view of what each other is doing or looking at. There are a number of studies addressing the effect of a shared workspace, and the general hypothesis is that sharing the same physical environment allows people to refer to the same set of objects and events and to coordinate conversational content, which leads to greater communication effectiveness and efficiency. Whittaker, Geelhoed, and Robinson (1993) studied the impact of adding a shared workspace on audio communication. The participants were given three tasks representing key features that are undertaken at a workplace: design layout, collaborative text editing and brainstorming task. The shared workspace was in the form of an electronic whiteboard which enabled people to view visual material jointly, to draw and to write. The results showed that participants in the audio plus shared workplace group obtained higher scores and took less number of turns in completing the task. Similar effects were observed in the collaborative editing task. Audio plus workplace enabled the direct depiction of spatial relationships for text edits and hence reduced the task completion time. In the contrary, there were few benefits of the shared workspace for the brainstorming task. The positive effect of shared workspace on collaborative tasks are supported by other studies (Kraut, Gergle, & Fussell, 2002; Olson, Olson, & Meadar, 1995).

In a more uni-directional communications study, Kraut, Miller, and Siegel (1996) investigated the effects of a shared visual environment in a bicycle repairing task. One expert and one novice formed a team to repair bicycle jointly. The novice had to make the repairs to the bicycle following the instructions provided by the expert through one of the following three methods: speech, speech and high quality video, one way audio and high quality video. The results revealed no significant differences in repairing qualities between the technology conditions, but there were differences in the process of giving advice. In the two video conditions, experts were more proactive when giving advice. Similarly, the result of Clark and Krych (2004) showed a non-significant difference in assembly error. However, the average building time was significantly (53.5%) shorter when the workspace was visible than when it was hidden.

Overall, there was strong evidence for supporting the effect of shared physical environment, in tasks that require physical and spatial manipulation, as the shared environment provides a straightforward way to jointly perceive and refer to objects.

Interactivity was defined as the ability to feedback via backchannels, completions, interruptions (Whittaker, 2002). Clack and Schaefer (1989) found that in interactive communication, feedback is often given by the listener concurrently with the speaker's utterance to indicate acceptance and understanding of what has been said. In addition, studies have showed that the ability of listeners to interrupt to clarify or question at the very moment when they experience communication breakdowns was an important mechanism for achieving moment by moment understanding (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Oviatt & Cohen, 1991).

Oviatt and Cohen (1991) studied the effect of interactivity on communication processes and communication efficiency. Ten pairs of expert-novice team collaborated to assemble a toy water pump. The experts were asked to give the novices a set of assembly instructions using two communication conditions: telephone and audiotape. In the telephone condition, the conversation happened in real time, allowing clarification exchanges and immediate feedback, whereas in the audiotape condition, the expert recorded the instructions and the novice listened to the tape with the freedom to play, stop and rewind. The study showed that novices in the telephone condition took significantly less time to complete the task. On the contrary, in the non-interactive condition, the communication was inefficient, for experts generated highly redundant instructions.

Similarly, Clark and Krych (2004) examined the performance difference between using interactive and non-interactive media. In the interactive media, the team communicated audibly. In the non-interactive condition, the director recorded the instructions blind to the builder, and builder later assembled their models from the recordings. There were 0.8% errors (7 out of 826) in the interactive condition, but 12.5% (74 out of 590) in the non-interactive condition, which was a significant difference.

Another paradigm to test the effect of interactivity was to investigate the negative impacts of preventing interactive feedback in speech communication. Kraut, Lewis, and Swezey (1982) manipulated the level of feedback that a listener can give to a speaker in a movie storyline summary task. They found that the amount of feedback that a speaker received during their summaries influenced the quality of listeners' comprehension. Communication was better the more feedback the communicator received.

Similarly, Bavelas, Coates and Johnson (2000) studied the relationship between listeners' feedback and speakers' performance. In their study, speakers told addressees an autobiographical story. Listeners' attention to the narrations was manipulated by instructing one half of the listeners to count the number of days until Christmas while speakers were telling the stories. Results showed that speakers with distracted listeners received less feedback and told stories less well, which confirmed that feedback influences narration quality. In particular, the endings of the narrations, which were intended to be suspenseful, fell flat when told to distracted addressees.

Kuhlen and Brennan (2010) studied the effect of two factors - speakers' expectations and addressees' feedback, on speakers' storytelling. They found that speakers' with attentive addressees who provided more feedback told the stories with more vivid details than those with distracted addressees only when they expected attentive addressees. Speakers with distracted addressees put less time into the task than those with attentive addressees only when they did not expect them to be distracted.

On the whole, interactivity is important in mediated communication. Adding interactive channels enable conversation parties to clarify doubts, suggest alternatives, and to jointly contribute to their understanding, resulting in decreased task completion time and increased comprehension.

Revisability/rehearsability was defined as the extent to which a medium enables the sender to rehearse or fine tune the message before sending (Clark & Brennan, 1991; Dennis & Valacich, 1999). Studies focusing on the effect of revisability/rehearsabilty on communication result are scarce and limited to its effect on the establishment of trust between communication participants. Walther, Slobacek and Tidewell (2001) suggested that the possibility that the amount of rehearsability that a communication system allows influences its users to create images of themselves or formulate them of others. A medium that has low rehearsability affords no opportunity to go back and reconsider the content, style, or diction of the communication. Rather, the contribution is transmitted to and accepted by the receiver as it is being formulated. The more rehearsability that exists, the more opportunity one has to reflect and focus on himself within a communication which in turn would impact the content of his messages (Joinson, 2001). Insight into the typing skills, grammar, or swiftness of communicants enable people to make judgments of ability, benevolence and integrity (Jarvenpaa, Knoll, & Leidner, 1998) and these judgments, in turn, are key in the development of trust.

Table **2-13** summarizes the effect of various affordances on communication effectiveness and efficiency. Technologies supporting text-based and redundant text-voice communication increase communication accuracy by allowing the participants to reprocess the messages, and those supporting speech communication shorten the time needed to complete a task. Technologies that transmit facial expressions and eye gazes benefit tasks where affect, emotion and attitude play a central role. Technologies that provide a shared visual environment are critical for tasks requiring complex reference to and joint manipulations of physical or spatial objects. Technologies supporting interactive feedbacks support communication effectiveness, and sometimes communication efficiency as well. Last, technologies that afford revisability show positive effect on the establishment of trust.

Table 2-13 Summary of the effects of various affordances on communication effectiveness and efficiency

Affordance		Findings (+ positive effect, - negative effect, * no significant effect, not reported)	
		Effectiveness	Efficiency
Reviewability / audibility	Text	+ positive (especially with increasing length of message)	- negative
	Speech	- negative (especially with increasing length of message)	+ positive
Redundant text-voice		+ positive	- negative
Visibility		+ positive for tasks with strong affect and emotion elements * no effect for tasks with strong cognitive elements	not reported /* no effect for tasks with strong cognitive elements
Shared workspace		+ positive for tasks with spatial cognition elements	+ positive for tasks with spatial cognition elements /not reported
Interactivity		+ positive	+ positive/not reported
Revisability		+ positive effect on trust	not reported

2.3.1.3 Media affordances relevant to clinical handovers

Hospitals adopt different handover protocols supported by various communication media. Although the amount of research on handover media is limited, some studies did suggest or imply that certain media affordances are important to ensure successful handovers.

Literatures from medical domains show a strong evidence for the positive effect of interactivity in handover communication. In clinical handovers, the Joint Commission has recommended that a proper handover protocol should include an opportunity for the participants to ask questions and clarify doubts (2006) and “Resilience” has been proposed to be one framing of handover functions (Patterson & Wears, 2010). Also, a few quasi-experimental studies and observational studies in medical handovers have supported that an interactive medium results in better handover results (Horwitz et al., 2009; Obstfelder & Moen, 2006; Randell et al., 2011).

Access to the same object of reference enables people to refer to the same set of objects and events and to coordinate conversational content, which leads to greater communication effectiveness and efficiency. In medical handover literature, there is a constant advocate for handover at bedside, where the handover participants can view patients as well as various medical documentations together. In addition, the study of Anderson and Mangino (2006) and Randell (2011) showcased the advantages of viewing patients and patients' medical records.

Medical literatures have suggested that the ability of seeing the handover participants results in better handovers outcomes. For example, Solet et al. (2005) directly argued that direct communication (e.g. face to face handovers, handovers at bedside) is preferred to mediated communication (e.g. handovers via telephone, email) in patient handovers. Also, being able to see communication parties may affect positively the "transfer of responsibility and authority (Patterson & Wears, 2010)". There is no experimental study focusing on the effect of visibility in medical handovers.

The effects of reviewability in medical handover has been examined through experimental studies (Bhabra et al., 2007; Pothier et al., 2005). These two studies clearly demonstrated the difference in efficacies of the handover styles. Pure verbal handover is not an effective handover method, while the typed handout is the most effective way of retaining information.

2.3.2 Theory of Task-Technology Fit

Section 2.3.1 examined the common ground theory and the effects of various affordances on communication effectiveness and efficiency. The examination showed that the effects of certain affordances were not universal. For example, visibility only benefited tasks where emotion plays a central role. Shared workspace only helped tasks involving a strong element of spatial cognition. Trying to account for these results, section 2.3.2 introduces the theory of task-technology fit and argues that in order for an affordance to benefit communication processes and outcomes, the “fit” between the affordance and the communication task should be taken into consideration.

2.3.2.1 Theory of task-technology fit

The concept of individual, task and technology fit dates back to studies on organizational theories, which address the relationship between human and task environment. For example, the theory of work adjustment defined correspondence as the “relationship in which the individual and the environment are corresponsive (Dawis, Lofquist, & Weiss, 1968)”. It states that the individual’s skills enable him to respond to the requirements of the work environment and the work environment respond to the requirements of the individual. Correspondence is the extent to which the individual fulfills the requirements of the work environment and the work environment fulfills the requirements of the individual. If the individual finds a corresponsive relationship between himself and the work environment, he will maintain it. If not, he tries to establish correspondence and if he fails, he will leave the work environment. Also, research on information processing and decision making proposed the concept of cognitive fit when studying the effect of information representation using either graphs or tables in supporting problem solving performance (Vessey & Galletta, 1991). It states that for most effective and efficient problem solving to occur, the information representation should support task requirements.

However, formal specification of task-technology fit (TTF) did not occur until the publication of three seminal papers in the area of information system (IS) research in mid 1990s (Goodhue, 1995; Goodhue & Thompson, 1995; Zigurs & Buckland, 1998). The underlying idea of TTF is that information system usage and performance benefits are attained through the fit between an information system and the task on which the technology is operated, as shown in Figure 2-8.

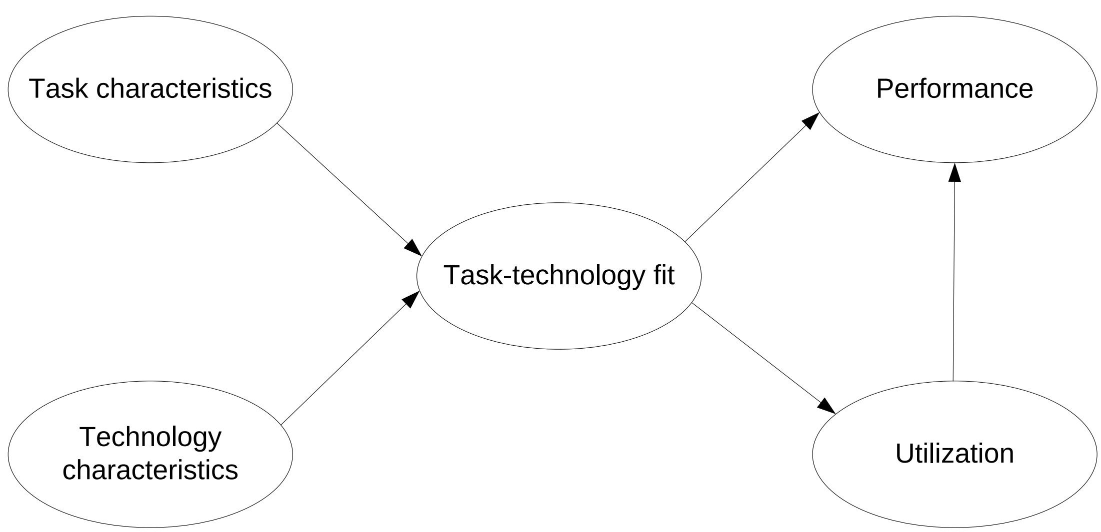


Figure 2-8 General model of task-technology fit (Source: Goodhue and Thompson, 1995)

Goodhue (1995) studied the link between task, technology, user and user evaluation of task-technology fit. The task dimension was accessed by measuring task routineness, and interdependence. The technology dimension was accessed by measuring the percentage of common systems, terminals of PCs per user, assisters per user and total assisters. Individual characteristic was measured by the computer literacy of users. 357 IS users from 10 large organizations participated in the survey study. The results showed that user evaluations of TTF were influenced directly by technology, task and individual characteristics. In addition, the interactions between technology and task were a significant predictor of user evaluation of TTF.

In another paper published in the same year, Goodhue and Thompson (1995) surveyed 600 users, varying from clerical to executive, of 25 different technologies, in 26 departments in two organizations, regarding the technologies they used in daily operations. Eight TTF dimensions were proposed for IS: quality of data, locatability of data, authorization for access to data, data compatibility, ease of use/training, production timeliness, system reliability, and relationship with users. Multiple regression analysis showed that task-technology fit, in particular the factors of data quality, timeliness, and IS relationship, directly predicts perceived utilization and perceived performance impacts.

Zigurs and Buckland (1998) expanded the theory of TTF to studies of Group Support System (GSS). They proposed four dimensions underlying GSS tasks: outcome multiplicity – the degree to which there is more than on desired outcome, solution scheme multiplicity – the degree to which there is more than one possible course of action to attain a goal, conflicting interdependence – the degree to which adopting one scheme conflicts with adopting another possible solution scheme, solution scheme/outcome uncertainty – the extent to which there is uncertainty about whether a given solution scheme will lead to a desired outcome. Based on the four dimensions, five types of task categories were presented as shown in Table 2-14.

Table 2-14 Aggregated GSS task categories (Source: Zigurs & Buckland, 1998)

	Simple tasks	Problem tasks	Decision Tasks	Judgment tasks	Fuzzy tasks
Outcome multiplicity	No	No	Yes	No	Yes
Solution scheme multiplicity	No	Yes	No	No	Yes
Conflicting interdependence	No	Yes or no	Yes or no	Yes or no	Yes or no
Solution scheme/outcome uncertainty	Not applicable	Low to high	Low to high	Low to high	Low to high

In addition, they proposed three factors for characterizing GSS technology: support for communication – the extent that a GSS supports, enhances or defines the capability of group members to communicate with each other, support for process structuring – the extent that a GSS supports, enhances, or defines the process by which groups interact, including capabilities for agenda setting, agenda enforcement, and facilitation, and support for information processing – the ability of a GSS to gather, share, aggregate, structure, or evaluate information. Based on the proposed task and technology dimensions, Zigurs and Buckland proposed the fit profiles as shown in Table **2-15**. The proposed fit profiles successfully predicted performance benefits of various GSS technologies.

Table 2-15 Fit profiles of task categories and technology dimensions (Source: Zigurs and Buckland, 1998)

	Communication support dimension	Process structuring dimension	Information processing dimension
Simple tasks	High	Low	Low
Problem tasks	Low	Low	High
Decision tasks	Low	High	High
Judgment tasks	High	Low	High
Fuzzy tasks	High	Medium	High

Though less emphasized in the seminal papers, the individual characteristics were proposed to be part of the construct. Goodhue & Thompson (Goodhue & Thompson, 1995) stated that "Perhaps a more accurate label for the construct would be task-individual-technology fit." The inclusion of individual characteristics is supported by the extension of TTF with fit-appropriation model (Dennis, Wixom, & Vandenberg, 2001) and with computer self-efficacy (Strong, Dishaw, & Bandy, 2006).

2.3.2.2 Task and Individual characteristics related to clinical handover

Task characteristics Previous literature review on technology mediated communication reveals the emotional-cognitive dimension of task characteristic. Some communication tasks require sophisticated social interaction but others do not. For tasks involving social interaction, communication technologies supporting the transmission of feelings, emotions, and attitudes led to better results. This suggests that communication media that support visible behaviors such as gazes, gestures, and facial expressions are preferred for this type of task. Although handovers serve many purposes including learning and colleague bonding, the primary objective of patient handovers is to ensure accurate transfer of information about a patient's state and plan of care (Patterson et al., 2004), Patient handover, therefore, is mainly a cognitive task rather than a socioemotional task.

Other than the emotional-cognitive dimension, task characteristics in communication have been studied from the perspective of information complexity. According to information theories, the most straightforward definition of complexity is the minimum possible length of a description of a model in some language (Casti, 1979; Crutchfield & Young, 1989). For example, Helleberg and Wickens (2003) varied the ATC instruction length that the pilot needed to read back and found that as the length of the ATC instructions increased, the portion of communication errors increased significantly across the text-based, the voice-based and the redundancy interfaces.

Patients have a great variability of complexities and care demand (Nemeth et al., 2008). There are patients who are recovering and will be discharged soon, as well as patients who are deteriorating and will be sent to the ICUs. Therefore, clinical handovers vary in information complexity. A task can be as simple as getting a patient's consent form, or as complex as operating on a sick patient. For a task with high information complexity, a larger size of description information is required to be handed over, thus the establishment of common grounds is more challenging.

Individual characteristics Studies have examined the effects of individual characteristics on general communication, including gender difference (Dennis, Kinney, & Hung, 1999), culture (Gudykunst & Kim, 1992), communication style (Norton, 1983) and communication participants' knowledge on communication topic (Lazarsfeld & Merton, 1964). Among these factors, communication participants' knowledge on communication topic was reported to be related to clinical handovers. Communication participants' knowledge on communication topic plays an important role in communication processes and results. It is well acknowledged that people communicate more easily with others of similar occupation and educational background, since they have similar experiences, beliefs, and knowledge (Lazarsfeld & Merton, 1964). Coiera (2000) stated that discussing a medical problem with a clinical colleague or with a patient leads to very different conversations. When communicating with a colleague, messages can be concise and much mutual knowledge can be assumed, whereas with a patient, more background knowledge needed to be explained to make the message understandable.

In clinical handovers, Cohen and Hilligoss (2009) summarized three broad categories of handovers frequently encountered in hospitals: a continuing patient transfer (a handover between professionals with similar expertise who are both familiar with the case), a new patient transfer (a handover between professionals with similar expertise who are not necessarily familiar with the case) and a cross-boundary transfer(a between professionals with distinctly different expertise). They stated that this classification was based on two factors that heavily influence handovers: the extent of shared knowledge, expertise and practice between handover participants, and the degree to which handover recipients know a patient to be handed over. This classification was supported by a few studies (Coiera, 2000; Nemeth et al., 2008).

A continuing patient transfer might occur when an on call team transfers a patient back to a primary team in charge. In this case, the handover participants already share a high level of common grounds of the patient. A relatively large amount of factual information could be assumed to be known to the recipient and the handover could focus on reporting changes and progresses of the patient over the on call period.

A new patient transfer occurs when the primary team transfers a patient to the on call team who has not seen the patient before. In such a handover, both parties have similar background and expertise. The recipient knows nothing about the patient and depends on the delivering party to create an accurate mental model for him.

A cross-boundary transfer occurs when an emergency department hands off a patient to an internal medicine department or when a tertiary hospital hands off a patient to a nursing home. In such cases, the handover participants have distinct knowledge and expertise and establishing common grounds becomes more challenging.

2.4 Research gaps

This chapter reviews the state of research on pre-handover and handover communication in healthcare. Subsequently, it covers theories and studies pertaining to doctors' information retrieval behavior during pre-handover and the effect of media during handover communication. The summary of research gaps is presented in Table **2-16**.

Table 2-16 Summary of research gaps

	Topic	Research gaps
Pre-handover	Pre-handover strategy and performance	• Doctors memory-intensive strategies in pre-handover imposed potential threats on patient safety; the underlying reasons for such strategies were unknown
	Information retrieval behavior	• Existing studies focused on the information access cost and none of them took into consideration the factor of 'performance' or 'accuracy', as suggested in the effort-performance space (Wickens et al., 2004) and the contingency decision model (Payne et al., 1988)
Handover communication	Communication media in clinical handover	• Existing studies were largely based on observations and expert opinions • Few experimental studies lacked theoretical foundations and had limited power of generalization
	Common ground theory	• Common ground theory fails to account for some research results. The concept of fit between individual, task, and technology is not well defined in common ground theory

The limited literature on doctors' pre-handover performance showed that they tended to use a memory-intensive strategy to retrieve patient information, by which they relied upon their memory rather than turned to written documentations such as patient medical records. Such memory-intensive strategies impose potential threats to patient safety, as human memory is vulnerable to errors (Wickens et al., 2013). However, there were no studies addressing the reason why doctors prefer memory-intensive strategies, and quantifying the potential threat imposed on patient safety.

Existing studies on information retrieval behaviors, primarily from the area of HCI have successfully illustrated that information access cost is an important factor in determining information access behaviors. People are sensitive to costs of low-level processes and adopt different interactive strategies when the costs change. As information access cost in the world increases, people adopt a more memory dependent strategy. However, none of the studies took into consideration the factor of ‘performance’ or ‘accuracy’, as suggested in the effort-performance space (Wickens et al., 2004) and the contingency decision model (Payne et al., 1988). According to them, one maximizes the utility of a choice by balancing the expected gain and the cost of effort. Likewise we posit that when retrieving knowledge from either the head or the world, people balance the expected gain of retrieving a desired memory item from the head or the world against the cost of memory search in the head or the cost of information access in the world (where accuracy is expected to be perfect). On this basis, the study of information access behavior would be incomplete unless the ‘expected performance/accuracy’ is taken into consideration.

Hence, this dissertation aimed to fill this research gap by investigating the underlying determinants for medical doctors’ memory-intensive strategies in the pre-handover phase and examining the effect of perceived accuracy and information access cost on information retrieval behavior and performance.

In the handover communication phase, there are only a limited number of studies addressing how communication media affect handover processes and results. Majority of the studies largely depend on observations in real setting and hence were not be able to examine causal relationships between a particular factor and handover quality. Only a few experimental studies explored the advantages and disadvantages of various handover methods (Bhabra et al., 2007; Pothier et al., 2005). However, these studies employed small sample size (5 subjects) and thus the validity of their results is questionable. Moreover, the experimental studies are not based on sound theoretical foundations. Therefore their results had limited power of generalization, though they were of high value to the particular settings being tested.

The review of common ground theory and technology mediated communication identified that the theory fails to take into account the task factor and the user factor. The concept of fit is not well developed in common ground theory and more often implied than defined. For example, studies showed that adding the affordance of visibility to audio communication only benefited tasks in which affect and emotion had a central role (Table **2-13**).

Furthermore, although the amount of research on handoff media is limited, some studies did suggest or imply that certain media affordances are important to ensure successful handoffs, including interactivity, reviewability, visibility and access to object of reference. (C. D. Anderson & Mangino, 2006; Horwitz et al., 2009; Joint Commission, 2006; Obstfelder & Moen, 2006; Solet et al., 2005). However, their effects were not validated empirically. Moreover, for the effect of visibility, expert clinician's recommendation was ironically in the opposite direction as predicted by empirical studies from technology mediated communication.

This research, therefore, aimed to expand the common ground theory by integrating the concept of TTF and to examine the effect of four affordances on handover communication effectiveness. Each of the affordances is shared in various numbers, by the six different communication methods labeling the rows (Table 2-17). In this manner, by comparing performance between various media, we will be able to identify the relative degrees of importance of the various affordance features. Moreover, based on the theoretical foundation, this research aimed to examine the effects of the four affordances through simulated experiments which would contribute to the existing literature with statistical evidence (Table 2-17).

Table 2-17 Manipulation of affordances and the corresponding handover methods

Affordance	**Interactivity** The extent to which handover sender and receiver can communication in real time.	**Reviewability** The extent to which handed over information retains over time.	**Shared workspace** Whether handover sender and receiver can view patient related information simultaneously.	**Visibility** Whether participants can see each other
Method A Receiver in Transit	Yes	Low	No	No
Method B Cell phone handover	Yes	Medium	No	No
Method C Written plus Verbal handover	Yes	High	No	No
Method D Audio Recording	No	Medium	No	No
Method E Face-to-face handover	Yes	Medium	No	Yes
Method F Distant consultancy	Yes	Medium	Yes	No

3. Conceptual framework and hypotheses

3.1 Conceptual Framework to study information retrieval in pre-handover and communication media in handover communication

A framework of clinical handover is proposed to model the factors influencing information retrieval in pre-handover and the factors affecting handover communication (Figure 3-1).

A handover begins with the handover sender collecting handover information from KIW or KIH. The information retrieval behavior and the pre-handover performance are determined by the information access cost and the confidence judgment (perceived accuracy), which in turn is affected by the feedback received from behavioral outcomes. Moreover, the optimal strategy is obtained by the cost-effective balance between information access cost and confidence judgment.

A handover communication occurs between a sender and a receiver, who possess intrinsic and extrinsic characteristics which can influence the communication processes and outcomes. Intrinsic characteristics include gender, culture, communication style and etc. and extrinsic characteristics include knowledge on a subject, familiarity with a patient and etc. The patient case is characterized by different dimensions as well, including complexity, structuredness, and emotion/cognition dimension. The sender transfers a patient case to the receiver through various types of communication media, which are characterized by their affordances. The goal of the handoff communication is for the sender and the receiver to establish a common ground of the patient. The common ground consists of two components: perception of elements in the patient case - perception of relevant patient information, and comprehension of the patient case -

integration of the perceived information to develop a comprehensive picture of the patient.

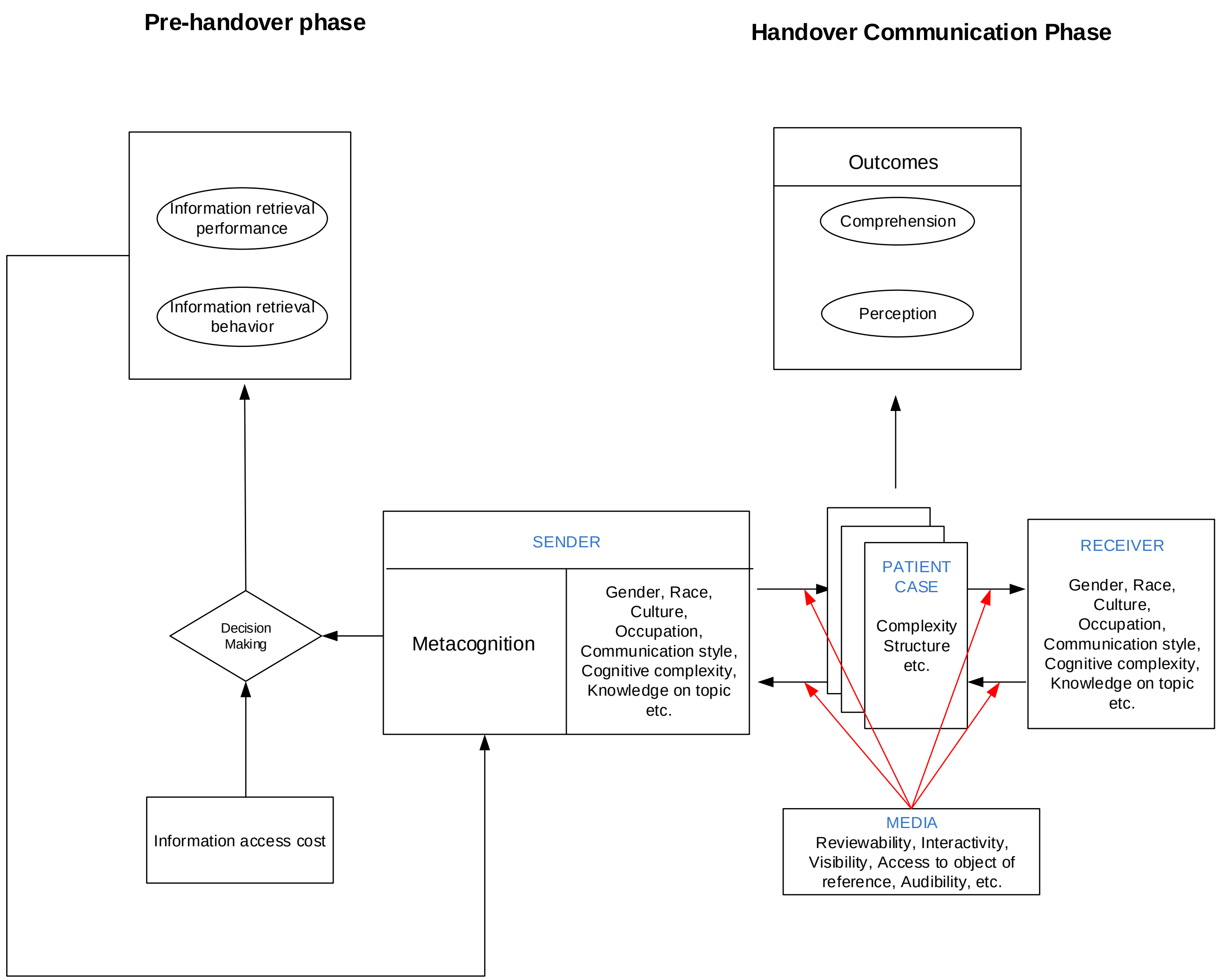


Figure 3-1 Conceptual framework for the pre-handover phase and the handover communication phase

3.2 Research model and hypotheses for studying information retrieval at pre-handover

Based on the conceptual framework, a reduced research model for studying information retrieval was drawn (Figure 3-2). This research argues that an information retrieval choice, between KIW and KIH is based, at least in part, on the balance between two forces. One influence on the choice between KIW and KIH is the **perceived accuracy** of the two techniques. If the information is safety critical, such as whether a patient is deathly allergic to a drug, it is essential that the information is retrieved from either the world, or the head with high accuracy. However, KIH may also be fairly accurate and physicians confident in such accuracy, for example, if the physician just read the chart, he may have a clear memory of the diagnosis or other patient data. In this case, the reason the physician may not rely on the near perfect KIW is because of the **cost**, in time and mental effort, if the KIW is not immediately available, for example, the chart has been relocated in a different room.

The influence of the two forces is complicated by one additional consideration. There is a growing body of knowledge that people are overconfident when judging their memory accuracy (Allwood et al., 2003; Bornstein & Zickafoose, 1999; Lichtenstein & Fischhoff, 1982). This dissertation argues that, overconfidence, rather than confidence judgment, affects pre-handover performance (the accuracy of information preparation). Under the condition of perfect calibration, where the confidence judgment reflects perfectly the actual accuracy, any changes in information access attempts due to varying confidence judgment can be considered as a result of adaptive decision making. For example, it is rational to make less access attempts to KIW if the confidence judgment, in this case, accuracy as well, increases from 70% to 80%. Under the condition of overconfidence, however, confidence judgment is overestimated, biasing the decision toward a KIH strategy more than it should and thereby impairing performance.

The focus of the research is on information access cost of KIW and the perceived accuracy of KIH. Information accuracy of KIW was assumed to be 100% and the cost of KIH is controlled to be constant. To simplify the notations, from this point onwards, IAC refers to access cost of KIW and confidence judgment refers to the perceived accuracy of KIH.

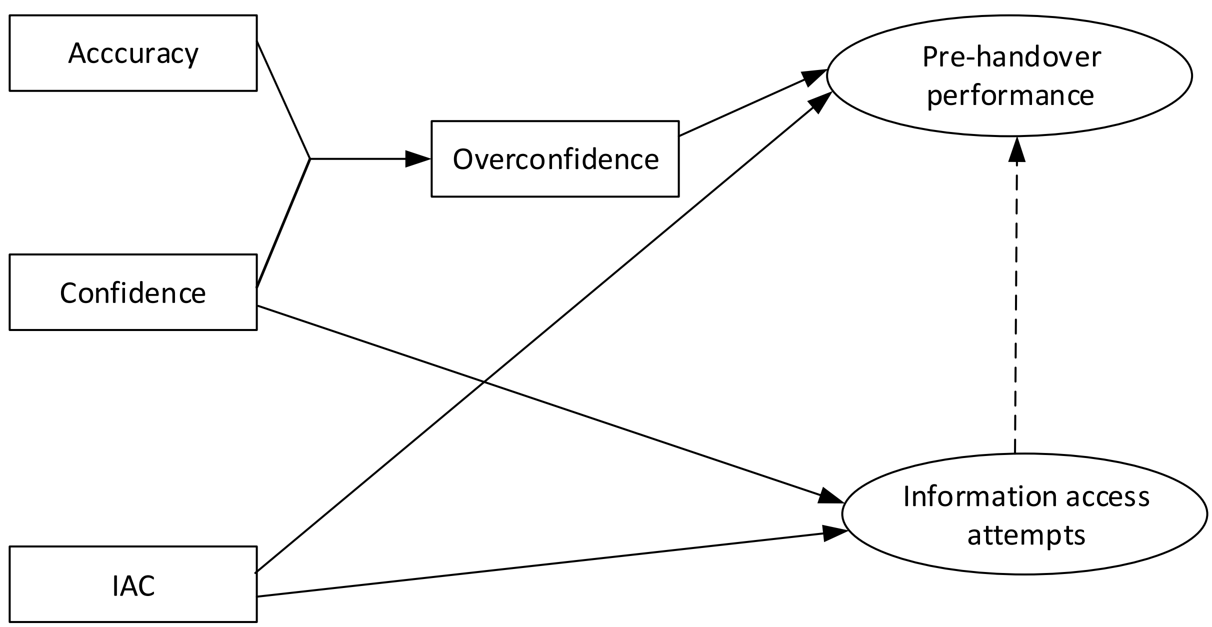


Figure 3-2 Effect of IAC, confidence, and overconfidence on information access attempts and pre-handover performance

Based on the research model, the following hypotheses were proposed.

Effect of IAC on information access attempts and performance Several studies showed that IAC affected people's information access behavior. As IAC increases, people adopt a more memory-intensive strategy and make fewer efforts to access information in the world. Further, the decrease in information access attempts to perfect KIW leads to poorer performance (Ballard et al., 1995; Gray & Fu, 2001, 2004; Gray et al., 2006).

As IAC increases, information access attempts will decrease and performance will decrease.

Effect of confidence judgment on information access attempt Several studies showed that when making decisions, the metacognitive judgment of accuracy, rather than the actual accuracy, determines the strategy selection (Chu & Spires, 2003; Fennema & Kleinmuntz, 1995; Payne et al., 1993). Therefore, the information access behavior is expected to be affected by the confidence judgment of memory accuracy, rather than the actual accuracy.

As confidence judgment increases, information access attempts will decrease.

Effect of overconfidence on performance Studies on metacognition have documented the prevalence of overconfidence bias (Fischhoff, Slovic, & Lichtenstein, 1977). Often, people are overconfident in judging their abilities, performance, and memory accuracy.

Under such conditions, people make fewer access attempts to perfect KIW than they should have done, resulting in compromised performance.

As overconfidence increases, performance will decrease.

Interaction between IAC and confidence judgment on information access attempts The effort-performance space suggests that when resources become scarce, the highest utility is achieved through allocating more effort to activities which yield the highest expected gain (Wickens et al., 2004). Under low IAC, additional information access attempts are still affordable even if confidence judgment is high. These unnecessary attempts can be considered as confirmatory feedbacks. When IAC is high, on the contrary, the incurred cost prohibits unnecessary references to KIW and impairs performance further.

There is an interaction between IAC and confidence judgment on information access attempts. As IAC increases, the effect of confidence judgment on information access attempts will be more prominent.

Interaction between IAC and overconfidence on performance Following the same rationale, under low IAC, the effect of overconfidence is moderated by the extra information access attempts, while under high IAC, the effect is more prominent.

There is an interaction between IAC and overconfidence on performance. As IAC increases, the effect of overconfidence on performance will be more prominent.

3.3 Research model and hypotheses for studying communication media

Based on the conceptual framework, a reduced research model for studying the effect of handover media was drawn (Figure 3-3). Although the amount of research on handoff media is limited, some studies did suggest or imply that certain media affordances are important to ensure successful handoffs, including interactivity, reviewability, visibility and access to object of reference (C. D. Anderson & Mangino, 2006; Horwitz et al., 2009; Joint Commission, 2006; Obstfelder & Moen, 2006; Solet et al., 2005). The four affordances are included in the research model.

Clinical handovers vary in information complexity, as patients have a great variability of complexities and have different level of care demand (Nemeth et al., 2008). There are patients who are recovering and will be discharged soon, as well as patients who are deteriorating and will be sent to the ICUs. For a task with high information complexity, more information is required to describe the state of the patient, thus the establishment of common grounds is more challenging.

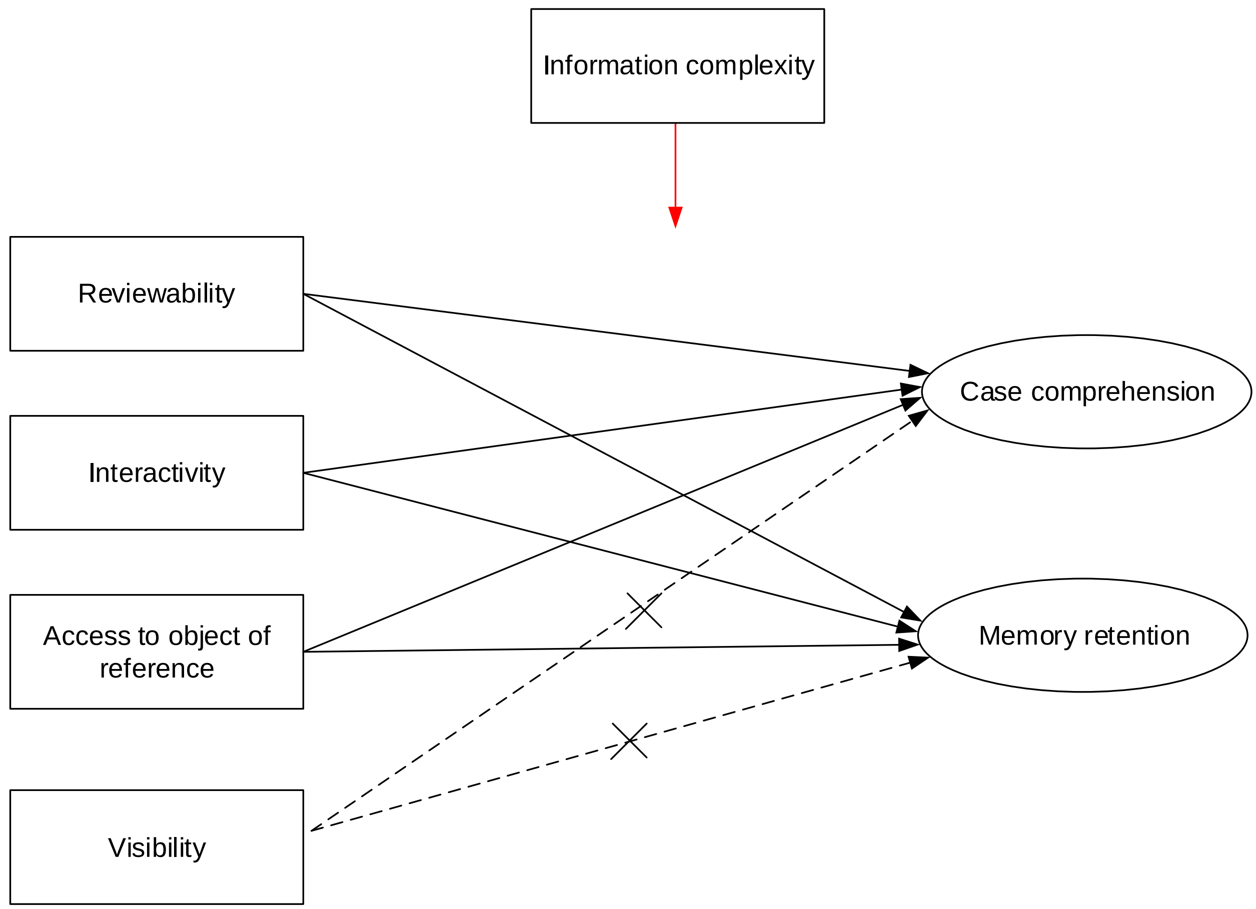


Figure 3-3 Effect of media affordances and task complexity on handover perception and comprehension

Based on the research model, the following hypotheses were proposed.

Effect of reviewability on handover perception and comprehension Reviewability is the extent to which a message can be reexamined or reprocessed within the context of a communication event and after the communication event ends (Clark & Brennan, 1991). With a medium supporting high level of reviewability, like text, handover participants can easily refer back and forth to any information during a handover. It supports conversational context and allows handoff participants to allocate more cognitive resources to interpret and synthesize information being handed over (Helleberg & Wickens, 2003; Lu et al., 2013; Wickens et al., 2011).

Communication medium supporting higher reviewability leads to better memory retention and case comprehension.

Interaction effect of reviewability and information complexity Tasks of higher information complexity involve the perception and comprehension of a greater amount of data and information, and it is more difficult to establish common ground. The effect of reviewability is expected to be more prominent when the information complexity is challenging working memory (Wickens & Gosney, 2003).

When information complexity is high, reviewability will more positively enhance memory retention and case comprehension

Effect of interactivity on handover perception and comprehension Being able to provide feedback and clarify doubts at the very moment when communication participants experience communication breakdowns has been shown to be an important mechanism for achieving shared understanding. Both parties jointly contribute to their understanding, which results in improved perception and comprehension (Bavelas et al., 2000; Clark & Krych, 2004; Kraut et al., 1982; Kuhlen & Brennan, 2010; Oviatt & Cohen, 1991).

Communication medium supporting higher interactivity leads to better memory retention and case comprehension.

Interaction effect of interactivity and information complexity The necessity to ask questions and clarity doubts is greater for complex tasks than simple ones. Thus the effect of interactivity will be more prominent for complex tasks.

When information complexity is high, interactivity will more positively enhance memory retention and case comprehension.

Effect of access to objects of reference on handover perception and comprehension Access to objects of reference for tasks involving reference and manipulation of physical objects enables communication participants to view the same set of objects they are discussing. It helps the senders and the receivers understand the current state of their task and enable them to communicate and ground their conversations more effectively (Clark & Krych, 2004; Kraut et al., 1996; Whittaker et al., 1993). In clinical handover literature, there is a constant advocacy for handover at bedside, where the handover participants can jointly view patients as well as various medical documentations together (C. D. Anderson & Mangino, 2006). The “fit” between technologies providing access to objects of reference (a shared workspace) and the clinical handover task is high.

Communication medium supporting access to object of reference leads to better memory retention and case comprehension.

Interaction effect of access to object of reference and information complexity Complex tasks usually involve a greater amount of data and information, whereas for simple tasks, it is easy to perceive and understand the handed over information and achieve a shared understanding. Thus the advantage of access to object of reference is expected to be more prominent for complex tasks.

When information complexity is high, access to object of reference will more positively enhance memory retention and case comprehension.

Null effect of visibility on handover perception and comprehension Visibility enables handover participants to see each other and express meanings using gazes, facial expressions and gestures. It helps transfer of feelings, emotions and attitudes (Stephenson et al., 1976; Weisband, 1992). However, clinical handovers, with the primary objective to transfer patient information and care plan accurately (Patterson et al., 2004), are characterized more as task-oriented cognitive tasks. Therefore, the "fit" between technologies affording visibility and the clinical handover task is low, and adding visual information to speech is not expected to contribute significantly to the establishment of common ground.

Communication medium supporting visibility does not lead to better memory retention and case comprehension.

4. Study 1: Exploratory study

The exploratory study aimed to investigate the handover practice in Singapore, with a specific focus on medical doctors' information retrieval strategies during pre-handover and the use of communication media during handover communication. The exploratory study would provide evidence for the proposed research models and guide the following three experimental studies

4.1 Research site

The study was conducted in the general medical ward of a Joint Commission International (JCI) accredited university hospital in Singapore. At the time of the study, the general medical service had 943 beds and 22 house officers (HOs), 53 medical officers (MOs), 30 registrars, and 85 consultants. When a patient was hospitalized, he/she was assigned to a primary care team comprising of 1 HO, 1 MO, 1 registrar and 1 consultant. The HO and the MO served as primary care providers. The registrar and the consultant functioned as supervisors.

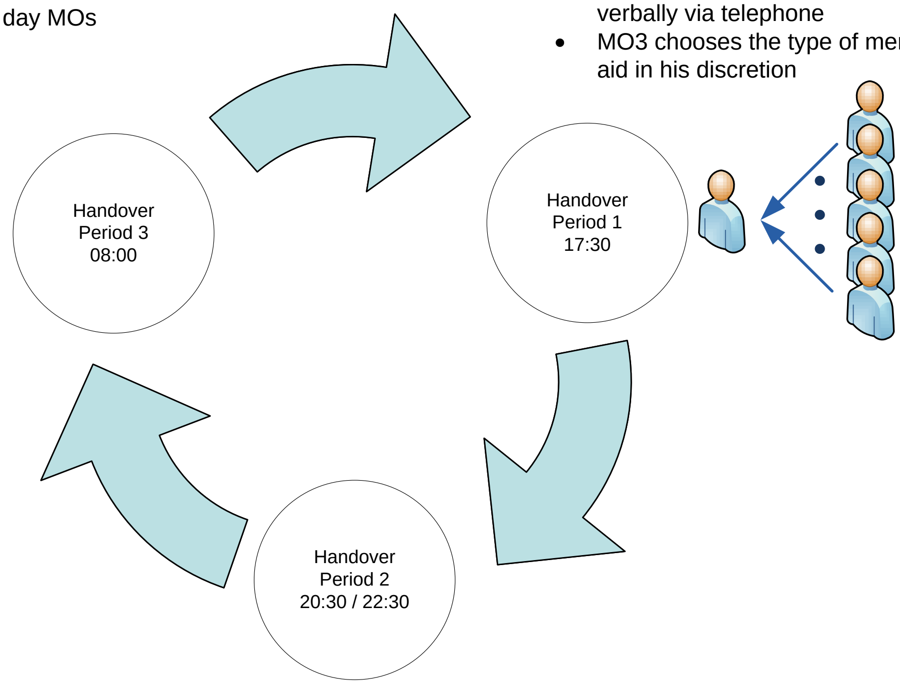
House Officers (HOs) are new medical graduates under clinical apprenticeship training (equivalent to "interns" in the US context). The training will typically last for a year and House Officers are required to perform 3 or more clinical postings. Medical Officers (MOs) are qualified medical doctors (MDs) who are immediately responsible for the total management of patients in the wards. They are required to mentor House Officers and help with the clerking of cases and performing of procedures. Registrars are experienced doctors with higher professional qualification and who are expected to undertake considerable responsibilities in addition to the duties of Medical Officers. In other words, they are considered as senior medical doctors (MOH Holdings, 2014).

The general medical ward ran a shift system, with three handovers occurring daily (Figure 4-1): night-day handover (0800), day-evening handover (1730) and evening-night handover (2030 /2230). Most of the time, handovers occurred between two MOs and occasionally between one registrar and one MO. Primary team physicians finished day work officially at 1730. After 1730, there were three evening on-call physicians (MO1, MO2 and MO3). Evening MO1 and MO2 were on active call, responsible for clerking and reviewing all newly admitted cases from 1730 till the end of their shifts. Evening MO3 was on passive call, responsible for reviewing all inpatients except the newly admitted. Evening MO1 and evening MO2 finished their shifts at 2030 and handed over to night MO1 and night MO2. Evening MO3 finished the shift at 2230 and handed over to night MO1 and night MO2. Night MO1 and MO2 were on both active and passive calls till 0800 the next day.

There was no well-established structure for handover in the general medical ward. Physicians could choose their own methods to hand over patients. A common method adopted by the physicians was phone handover. Outgoing physicians made phone calls to incoming physicians and incoming physicians wrote down critical information.

- Night on call medical officers handover patient information to corresponding day MOs

- Outgoing medical officers handover patient information to evening on call medical officer (MO3)
- Patient information exchanged verbally via telephone
- MO3 chooses the type of memory aid in his discretion



- Evening on call medical officer (MO3) handover patent information to night medical officers (night MO1 and night MO2)
- Night MO1 and night MO2 choose the type of memory aid in his discretion

Figure 4-1 Shift cycle at general medical ward

4.2 Methods

4.2.1 Participants

The National Healthcare Group (NHG) Domain-Specific Review Boards (DSRBs) approved this study and all participants gave their informed consent. Twenty-three MOs (14 female and 9 male) participated in the study. All participants were undergoing training in different specialties, with experience ranging from 16 months to 5 years (mean = 3.5 years, SD = 1.2 years). Their ages ranged from 24 to 33 (mean = 27.3 years, SD = 2.0 years). Sixteen participants graduated from medical schools in Singapore, 4 from the United Kingdom, and 3 from elsewhere.

4.2.2 Data collection

Semi-structured interview and shadowing were used in the exploratory study. Semi-structured interview affords two-way communication and allows interviewers to probe follow up questions. Shadowing is an ethnographic research method, which involves a researcher closely following a member of an organization over an extended period of time. It provides researchers with a rich, comprehensive and systematic picture of exactly what people do at work (McDonald, 2005).

Table **4-1** shows the questions guiding the semi-structured interview.

Table 4-1 Semi-structured interview questions

When you are about to finish your shift as a day medical officer
• Please describe your normal practice of handing off patient information to MO3 • How do you decide who to hand off? (classify cases to hand off) • How do you collect all the relevant information? • How do you make sure the requested task has been completed?
When you are on-call MO3

• How do you remember the information handed off to you? • If you are based on written notes, how do you structure your information? • How do you prioritize the importance of patients? • Please describe the type of tasks you are requested to perform during your duty as MO3.
Opinions and suggestions
• Can you identify weaknesses/strength in your handoff practices? • Do you believe the current handoff practices could be improved? If yes, how?

The data was collected on weekdays over a three-month period. The semi-structured interview was conducted on the afternoon of the day when a participant was on evening call (evening MO3) and shadowing of the participant was conducted from 17:30 to 19:00 on the same day. The participant was provided with a smart phone, through which handover conversations were recorded and time stamped. After receiving each handover call, the participant was requested to rate the handover quality on a 5-point Likert scale (1= very poor, 5 = excellent). In addition, the participant was asked to share their handwritten notes. The submission of the post-handover ratings and the handwritten notes was voluntary and the participant could opt out for any reason. In total, 96 handover calls covering 152 patient cases were collected. Post-handover ratings were collected for 98 patient cases (64.5%) and handwritten notes were collected for 102 patient cases (67.1%).

4.2.3 Data analysis

Interview data and verbal handover communication were transcribed using Transcriber® (http://trans.sourceforge.net/). A qualitative analysis of the interview data was conducted, following standard procedures for inductive analysis (Johnson & Christensen, 2004), where two human factors researchers reviewed the data independently and coded the themes. Differences in the coding were discussed and negotiated until consensus was reached.

To characterize information transfer during handovers, a coding scheme was developed to categorize patient information. First, five coding schemes published recently were analysed (Alem et al., 2008; Flanagan, Patterson, Frankel, & Doebbeling, 2009; Sledd, Bass, Borowitz, & Waggoner-Fountain, 2006; Solet et al., 2005; Ye et al., 2007) and then synthesized by merging similar information categories. Second, a panel of Subject Matter Experts consisting of 1 consultant and 1 registrar reviewed and refined the coding scheme. Table 4-2 shows the final coding scheme comprising 10 main categories and 29 sub categories.

Table 4-2 Handover Information Coding Scheme

Main Category	Sub-category
Physician identity	• MO subspecialty • MO name • Registrar name • Consultant name
Patient identity	• Ward/bed • Patient name • Patient age • Patient race • Patient gender • Patient ID
Medical history	• Pre-morbid status • Past medical history including long term illness • Allergy
Diagnosis	• Time of admission • Reason for hospitalization/ chief complaint • primary / working diagnosis including hypothetical diagnosis
Current conditions	• Patient progress and current condition • Active problem list: patients' active issues which lead to the handover
Test & consultation	• Diagnostic tests performed and results during hospitalization • Pending tests • Consultation performed from sub-specialty or other departments and results • Pending consultation / referral letter
Medication	• Active medications / treatment
Primary team management plan	• Short-term plan (plans for the next 6-24 hours) • Long-term plan • Code status
Cross-covering task	• Explicitly "to do" list for cross-covering doctors • 'What-if' plans for cross-covering doctors if any abnormality occurs
Psychosocial concerns	• Psychosocial concerns

The registrar trained one human factors researcher on how to code patient information. One information category was considered as discussed if the participant made some comments belonging to that category. For example, "No, we didn't do a CT brain (scan)" was regarded as a discussion of 'diagnostic test', even though the test was not performed. Before proceeding to the full coding, the reliability of the coding scheme was verified by checking the inter-rater reliability between the registrar and the human factors researcher. Both of them independently coded 30 randomly selected transcripts. Cohen's kappa was calculated (k = 0.852), which showed good agreement between the two. The human factors researcher then continued to code the remaining communication transcripts and the handwritten notes. An example of a communication transcript and corresponding information coding are shown in Appendix B.

4.3 Results

4.3.1 Interview results

The number of patients that a primary team physician was responsible for ranged from 5 to 20, depending on the subspecialty from where the physician came. Normally, a primary team physician handed over less than three patients. These included sick patients, patients with important pending diagnostic test results, patients with pending referrals from other subspecialties, patients with special code statuses, patients with demanding families, and patients with behavioral or mental problems. Participants also commented that the handover decisions were team based decisions, which were made together with the consultants and the registrars at the exit rounds.

Several information sources were referred to in order to collect handover information, including the computerized medical system, patient case notes, patient lists and personal notes. However, a majority of the participants (18/23) stated that they usually did not collect patients' information from different information sources. Instead, they retrieved patient information directly from memory. This was illustrated by one participant's comments on how she collected patient information: "Generally, because I know the patient very well, I just think about it for a few seconds, and try and make some coherence before I tell MO3 (the evening on-call physician). Sometimes I write it down, but usually I just think about it."

The participants reported that as primary team physicians, they checked the computerized medical system and the patient case notes the next day to ensure that the requested tasks were completed.

All the participants stated that when working as on-call physicians, they took notes during handovers. A majority organized their notes according to chronological sequence of handover calls, while few physicians organized according to task type or patient location. Most of the time, the primary team physicians did not indicate the severity of patients during handovers. Rather, the on-call physicians decided and prioritized the importance of patients.

Four major types of 'to do' tasks during the on-call period were mentioned: reviewing sick patients, tracing diagnostic test results, tracing referral reply, and attending to emergencies. All participants stated that they gave priorities to reviewing sick patients and attending to emergencies over tracing diagnostic test results and tracing referral replies.

Eight participants shared their view on the strengths of the current handover practice, including symmetric communication (1/8), convenience of communication obtained by the mobility of phone calls (5/8), and affordance of clarification obtained by verbal communication (3/8).

Seventeen participants identified three main areas of weaknesses for the current practice. The first area was concerned with communication content, including omission of important information (13/17), disorganization of handover information (3/17) and information overload (2/17). The second area of weaknesses focused on handover media. Difficulty to convey information over the phone (3/17) and lack of official written documents (4/17) were identified as the two major pitfalls. Difficulty to convey information was illustrated by one participant's response - "There may be miscommunication over the phone, but I try to eliminate that by getting the person to repeat, or if I answer the phone call, I repeat to the person exactly what I am supposed to do. Also there is no proper documentation of the handover per se." The third area targeted the organizational factors, including high workload (2/17) and no protected time for handover calls (1/17).

The participants identified several possibilities for improving the current handover practice, including enforcement of mandatory information categories such as double identifier (i.e. two non-modifiable identifiers are needed to identify a patient, such as patient ID and patient name.), 'to do' tasks and contingency plans, and providing written document besides verbal communication.

4.3.2 Shadowing and post-handover rating results

Table 4-3 shows the occurrence and percentage of each information category. 'Information exchanged in a handover' counted the information mentioned in a handover. 'Information voluntarily delivered by sender' was the information transferred voluntarily by a primary team physician to an on-call physician. 'Information prompted by receiver' counted the information prompted by an on-call physician, if not delivered voluntarily by a primary team physician. 'Information written down by receiver' counted the information category written by an on-call physician on the handwritten note.

Table 4-3 Information transfer between primary team physicians and on-call physicians for 152 patients (n=number of occurrence)

Main Category	Sub-category	Information exchanged in a handover *n (%)*	Information voluntarily delivered by sender *n (%)*	Information prompted by receiver *n (%)*	Information written down by receiver *n (%)*
Physician Identity	MO subspecialty	20 (13)	20 (13)	0 (0)	1 (1)
	MO name	81 (53)	81 (53)	0 (0)	1 (1)
	Registrar name	7 (5)	4 (3)	2 (1)	1 (1)
	Consultant name	1 (1)	1 (1)	0 (0)	0 (0)
Patient Identity	Ward/bed	152 (100)	145 (95)	9 (6)	102 (100)
	Patient ID	12 (8)	9 (6)	3 (2)	3 (3)
	Patient name	20 (13)	16 (11)	4 (3)	3 (3)
	Patient's age	56 (37)	54 (36)	2 (1)	20 (20)
	Patient's race	17 (11)	17 (11)	0 (0)	5 (5)
	Patient gender	102 (67)	102 (67)	0 (0)	20 (20)
Medical history	Pre-morbid	11 (7)	10 (7)	1 (1)	1 (1)
	Past medical history	66 (43)	63 (41)	3 (2)	35 (34)
	Allergies	4 (3)	3 (2)	1 (1)	2 (2)
Diagnosis	Time of admission	23 (15)	22 (14)	1 (1)	1 (1)
	Reason for hospitalization	73 (48)	72 (47)	3 (2)	38 (37)
	Primary / working diagnosis	114 (75)	114 (75)	8 (5)	56 (55)
Current condition	Patient progress /current condition	89 (59)	74 (49)	33 (22)	24 (24)
	Active problem	73 (48)	73 (48)	3 (2)	26 (26)

	list				
Test & consultation	Diagnostic test performed	75 (49)	69 (45)	16 (11)	24 (23)
	Pending test	73 (48)	72 (47)	11 (7)	28 (27)
	Consultation performed	9 (6)	7 (5)	2 (1)	2 (2)
	Pending consultation	29 (19)	29 (19)	1 (1)	10 (10)
Medication	Active medication	82 (54)	72 (47)	29 (19)	26 (25)
Primary team management plan	Short-term plan	50 (33)	46 (30)	6 (4)	16 (16)
	Long-term plan	1 (1)	1 (1)	0 (0)	0 (0)
	Code status	25 (16)	19 (13)	7 (5)	9 (9)
Cross-covering task	To-do	138 (91)	131 (86)	14 (10)	81 (80)
	What-if plan	75 (49)	66 (43)	21 (14)	29 (28)
Psychosocial concerns	Psychosocial concerns	11 (7)	10 (7)	1 (1)	0 (0)

* The sum of information voluntarily delivered by sender and information prompted by receiver did not necessarily equal to information exchanged in a handover. The outgoing physicians can deliver information under one category and the on-call physicians can request for more details under the same category.

Only one information category - *ward/bed*, as an identifier of the patient, was exchanged for every handover case. Other categories exchanged frequently were *to-do task* (91%), *primary/working diagnosis* (75%), *patient gender* (67%), *patient progress and current condition* (59%), *active medication* (54%), *MO name* (53%), *what-if plan* (49%) and *diagnostic test results* (49%).

The primary team physicians initiated the majority of the discussions. Five categories were voluntarily delivered in more than half of the cases: *ward/bed* (95%), *to do task* (86%), *primary/working diagnosis* (75%), *patient gender* (67%), and *MO name* (53%). Information prompted frequently by the on-call physicians included *patient progress and current condition* (22%), *active medication* (19%), *what-if plans* (14%), *diagnostic test performed* (11%), and *to do task* (10%)

The on-call physicians wrote down only one information item, *ward/bed* (100%) for every patient. Other information items written down frequently included *to do task* (80%), *primary/working diagnosis* (55%), *reason for hospitalization* (37%), *past medical histor*y (34%), *pending test* (27%), *active problem list* (26%), *active medication* (25%) and *patient progress and current condition* (24%).

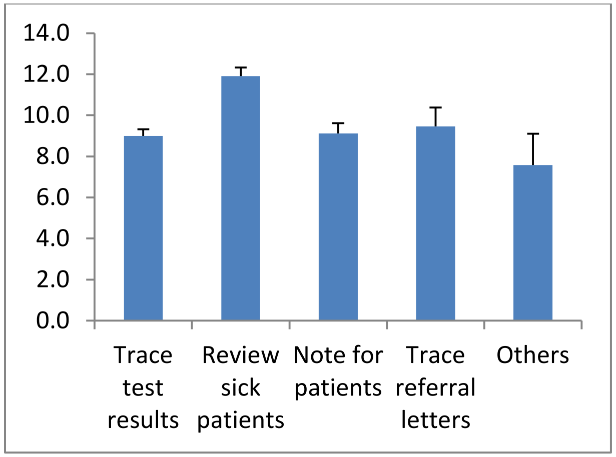
An in-depth analysis of the to-do tasks showed that there are five major types of tasks (The 13 cases without explicit statement of the to-do tasks were classified based on other information discussed in the verbal handover communication): *review sick patients* (42 patient cases, 28%), *trace diagnostic test results* (71, 47%), *trace referral letter replies* (i.e. a primary team physician refers a patient to another physician) (15, 10%), *note for patients* (i.e. a primary team physician informs an on-call physician of patients with high possibility of deteriorating during the on-call period) (17, 11%). *Others* included tasks related to administrative matters such as, obtaining a signature from a patient (7, 5%).

Ppostulating that information transfer, handover duration and handover quality may differ across task types, the number of information categories exchanged in a handover, voluntarily delivered by sender, prompted by recipient, written down by recipient, the handover duration and the post-handover rating were tabulated for each patient case (Table 4-4). The average rate of exchange was 9.8 categories of information per handover. The primary team physicians voluntarily delivered 9.2 categories of information. The on-call physicians prompted for 1.2 categories of information and recorded 5.5 categories of information on their handwritten notes. Handover of a patient took a mean of 80 seconds. The majority of handover qualities were perceived as good, with a mean score of 3.6 (1= very poor, 5 = excellent).

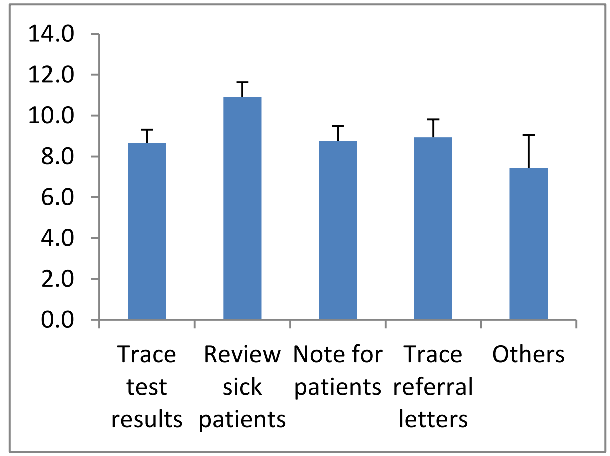
Table 4-4 Information transfer, duration and perceived handover quality of each type of task (Mean (SD); n=number of handover cases)

	All	**Review sick patients**	**Trace diagnostic tests**	**Trace referral letters**	**Note for patients**	**Others**
Information exchanged (n=152)	9.8 (3.2)	11.9 (2.7)	9.0 (2.8)	9.5 (3.6)	9.1 (2.1)	7.6 (4.4)
Information voluntarily delivered (n=152)	9.2 (3.0)	10.9 (2.8)	8.6 (2.7)	8.8 (3.1)	8.8 (1.9)	7.4 (4.3)
Information prompted (n=152)	1.2 (1.3)	2.0 (1.4)	1.0 (1.1)	0.8 (0.9)	0.7 (0.8)	0.3 (0.5)
Information written down (n=102)	5.5 (2.1)	6.6 (2.2)	5.1 (1.9)	5.8 (2.0)	5.3 (2.2)	3.2 (1.5)
Duration for each case (sec)	80 (51)	106 (53)	68 (43)	75 (62)	78 (53)	50 (27)
Handover quality	3.6 (0.8)	3.7 (0.8)	3.4 (0.9)	3.9 (1.0)	4.0 (0.0)	4.3 (0.6)

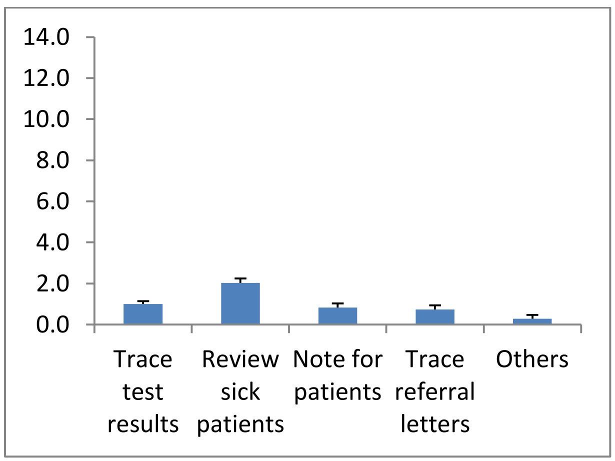
Kruskal-Wallis tests were conducted to examine the effect of task type on the four measures of information transfer and handover quality. The results revealed significant effects of task type on information exchanged in a handover ($F(4,147) = 26.21$, Asymp.Sig. <.001), voluntarily delivered by sender ($F(4,147) = 20.54$, Asymp. Sig. <.001), prompted by on-call physician ($F(4,147) = 21.58$, Asymp.Sig. <.001) and written down by on-call physician ($F(4,97) = 17.02$, Asymp.Sig. <.01) (Figure 4-2). 'Review sick patients' was the highest in all four measures of information transfer and 'others' was the lowest. In addition, there was no significant difference in perceived handoff quality between different task types ($F(4,93) = 8.213$, Asymp.Sig. = .125). A one-way ANOVA was conducted to examine the effect of task type on handover duration. The results revealed a significant difference in handoff duration ($F(4, 147) = 4.916$, $p < .001$), with ‘review sick patient’ taking the longest and ‘others’ the shortest.



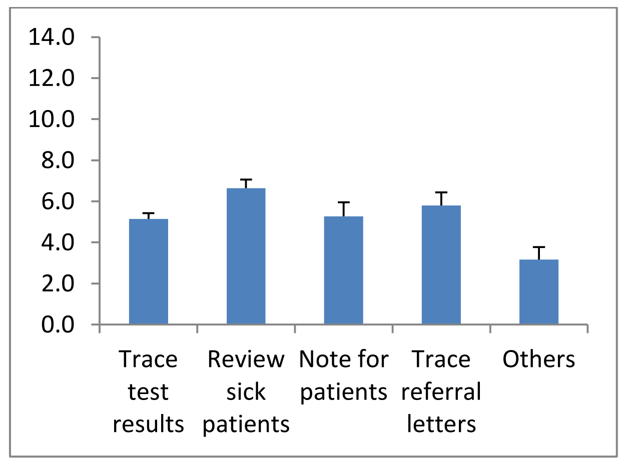
(a) Information exchanged (Sig. <.001)



(b) Information voluntarily delivered by sender (Sig. <.001)



(c) Information prompted by receiver (Sig. <.001)



(d) Information written down by receiver (Sig. <.01)

Figure 4-2 Information transfer during handovers

4.4 Discussion

Consistent with existing literature (Alem et al., 2008; Vidyarthi et al., 2006), results from the exploratory study showed that during pre-handover physicians tended to rely on their memory for patient information rather than refer to written documents. The inclination towards memory-intensive strategies indicated that pre-handover performance largely depends on primary team physicians' ability of retrieving patient information from memory accurately. However, knowing information is stored in long-term memory does not guarantee that it will be retrieved successfully when needed (Wickens & Hollands, 1999). Moreover, a considerable amount of studies have shown that people are overconfident in eyewitness memory and memory for general knowledge (Bornstein & Zickafoose, 1999). Failure to retrieval or inaccurate retrieval of any critical patient information may lead to suboptimal handover preparation, resulting in near misses or adverse events.

The results provided evidence for the importance of interactivity in patient handovers. The affordance of clarification obtained by verbal communication was identified by the physicians as a major benefit of the current handover practice. In addition, the analysis of information transfer between outgoing physicians and on-call physicians showed that although outgoing physicians initiated the majority of the discussion, on-call physicians did prompt for more patient information. The outgoing physicians and the on-call physicians worked together to achieve a shared understanding of a patient.

The practice of mobile phone handover supported the affordance of semi-reviewability, where the handover receivers were allowed to record information at their own discretion. The results showed that even under the affordance of semi-reviewability, there was large information loss in the process. On average, 9.8 categories of information were discussed in a handoff but only 5.5 were recorded on paper by the on-call physicians (Table 4-4). Moreover, there were large disparities between information exchanged and information written down, even for categories bearing critical patient information such as *what-if plan* (Table 4-3). The information loss may have a strong negative impact on handover communication quality and patient safety. According to previous studies, the memory retention of patient information using verbal only handover was extremely low (Bhabra et al., 2007; Pothier et al., 2005). In the current practice, therefore, it was highly unlikely for the on-call physicians to recall information which had not been recorded down. Nearly 40% of handover information was lost from the outgoing physicians to the on-call physicians. As summarized in 'human factors of transition of care', the on-call physicians are experiencing high mental workload during handover communication (Harvey, Schuster, Durso, Matthews, & Surabattula, 2006). Their activities include perceiving and understanding the transferred information, probing for more details when necessary, and synthesizing all the information to create an accurate mental model of the patient. Under such high mental workload, it is not surprising to find high amount of information loss.

Results from the exploratory study also provided statistical evidence that information complexity varied significantly across patient cases. There are significant differences in information exchanged, voluntarily delivered by sender, prompted by recipient and written down by recipient. The difference in information complexity was further supported by the varying handover duration across task types. The results were consistent with the SME panel's opinion that review sick patients was usually of high severity, and should be discussed in more detail over the other four types of tasks.

To summarize, the exploratory study provided qualitative evidence for the memory-intensive strategies when handover sender collected patient information at the pre-handover phase. To examine the determinants of such strategies, Study 2 and Study 3 were designed and conducted. Moreover, it offered qualitative and quantitative support for the effects of interactivity, reviewability, and task complexity on handover communication Based on the results of the exploratory study, a simulation-based experiment (Study 4) was carried out to examine the effects of the four affordances, and their interaction effects with task complexity.

The exploration study was subject to the limitation of the handover practice employed by the department. Phone handover only afforded interactivity and semi-reviewability, thus it was impossible to investigate the affordance of visibility and access to subject of reference. The takeaway points of Study 1 are summarized as follows.

- During pre-handover, handover senders tended to rely on their memory for accessing patient information
- Two-way handover communication allows handover receivers to seek further information and clarify doubts
- There are large information loss even during semi-reviewability condition

- Information complexity during handover communication varies according to the severity of patients.

5. Study 2: Determinants of information retrieval strategy

Study 2 aimed to examine the factors influencing information access behavior and performance in general. Specifically, as indicated in Figure 5.1 (Research model between picture type, IAC, confidence judgment, information access attempts and performance), study 2 tested 6 hypotheses. Due to its nature, confidence judgment and overconfidence cannot be directly manipulated as independent variables. Therefore, they were “manipulated” using the Tulving recognition task (1981), which was known to induce an accuracy-confidence inversion between two types of picture pairs. Picture type (AA’ vs. BC’) was the first independent variable. The second independent variable - IAC from the world, was manipulated by imposing varying delay time when information in the world was accessed. There were two dependent variables: the number of information access attempts and the recognition task performance. Six hypotheses are formed and listed as follows (**Figure 5-1**).

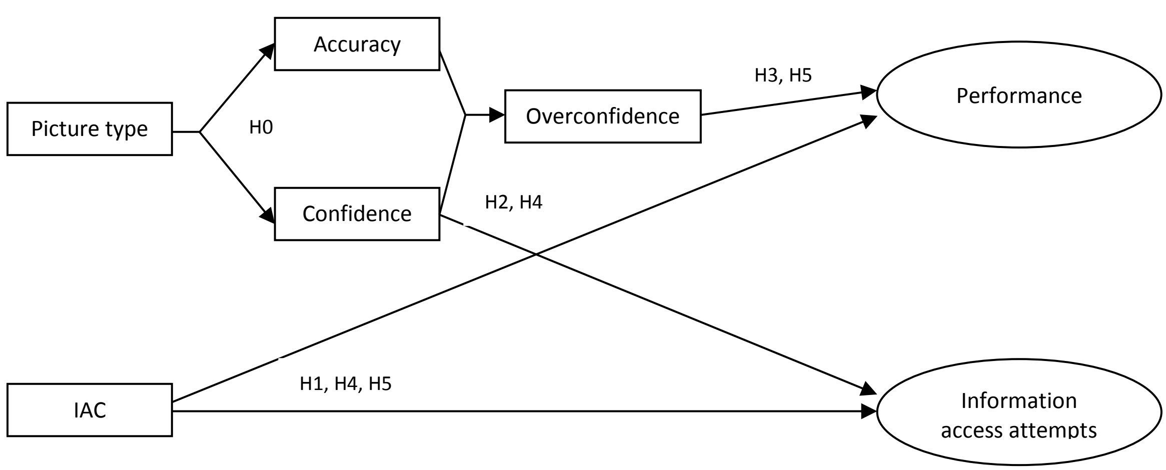


Figure 5-1 Research model between picture type, IAC, confidence, information access attempts and performance

H1: There is an inverse pattern of accuracy and confidence judgments between picture type AA' and BC'. Picture type AA' will produce higher memory accuracy but lower confidence, resulting in lower overconfidence, while picture type BC' will produce lower memory accuracy but higher confidence, resulting in higher overconfidence.

H2: As IAC increases, the number of information access attempts in the world will decrease and the recognition performance will decrease.

H3: As confidence judgment increases, the number of information access attempts in the world will decrease.

H4: As overconfidence increases, recognition performance will decrease.

H5: There is an interaction between IAC and confidence judgment on information access attempts. As IAC increases, the effect of confidence judgment on information access attempts will be more prominent.

H6: There is an interaction between IAC and overconfidence on recognition performance. As IAC increases, the effect of overconfidence will be more prominent.

5.1 Method

5.1.1 Stimuli

1000 color photos with resolution of 1024 ×768 were downloaded randomly from National Geographic Photo of the Day. These photos covered shots of animals, nature and weather, landscapes, people and culture, and travel. From the initial pool of 1000 color photos, a set of 180 photos was selected as "test pictures" on the basis of similarity ratings of the halves of each bisected photo. Similarity ratings were assessed on a scale of 1 (not similar) to 5 (perfectly similar) by 6 participants not involved in the experiment. The "test pictures" selected ranged in similarity from 3.4 to 4.0. Another set of 180 photos with similarity ratings from 1.8 to 2.8 was selected as "buffer item pictures". One half of each photo was randomly selected as the target picture (A, B, C…) and the other half its similar counterpart (A', B', C'…).

5.1.2 Participants

Thirty-nine undergraduate students of both genders at Nanyang Technological University participated in the experiment. Each subject was paid 20 Singapore dollars (16.4 USD) for their participation. The results of 3 participants were excluded from the analysis because their recognition rate in the recognition test w/o aid (test 1) was below 40%.

5.1.3 Experimental task

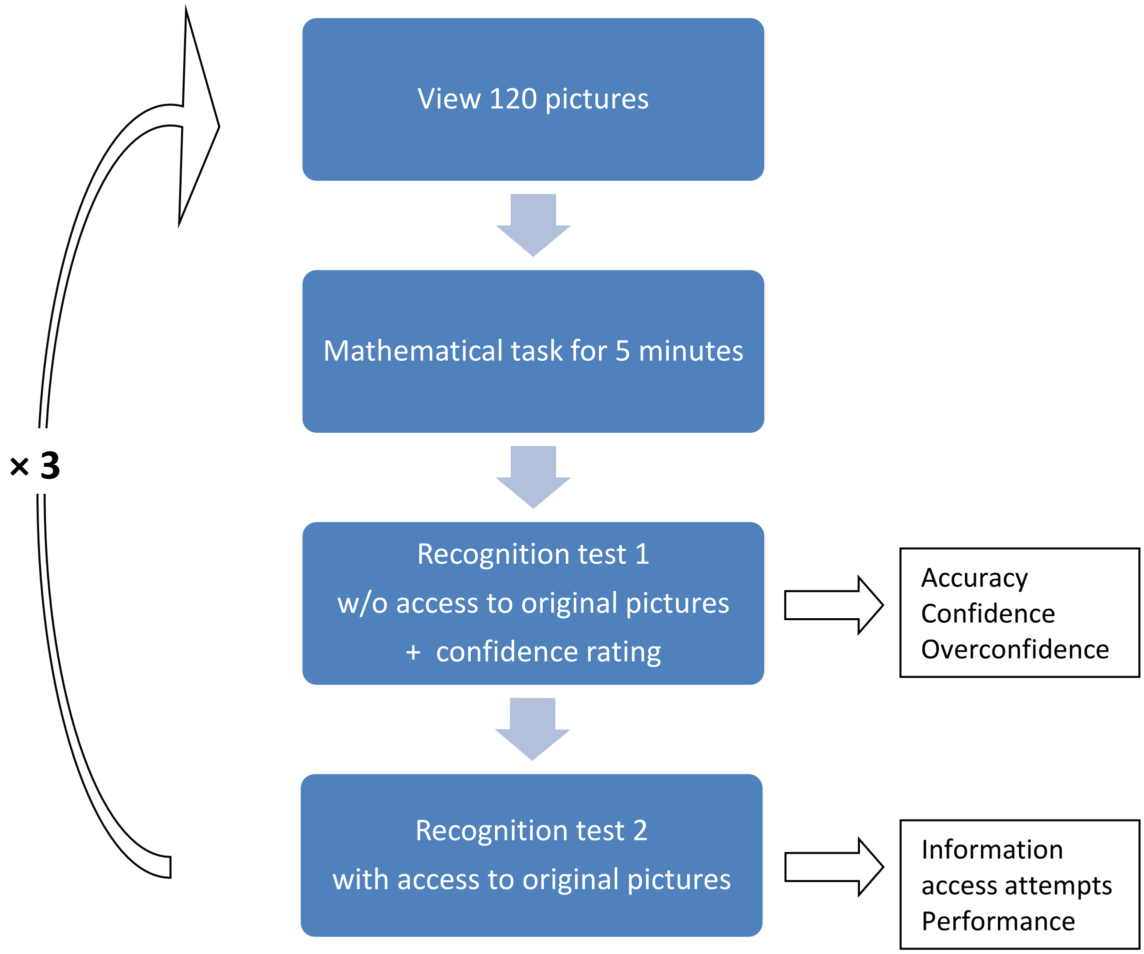


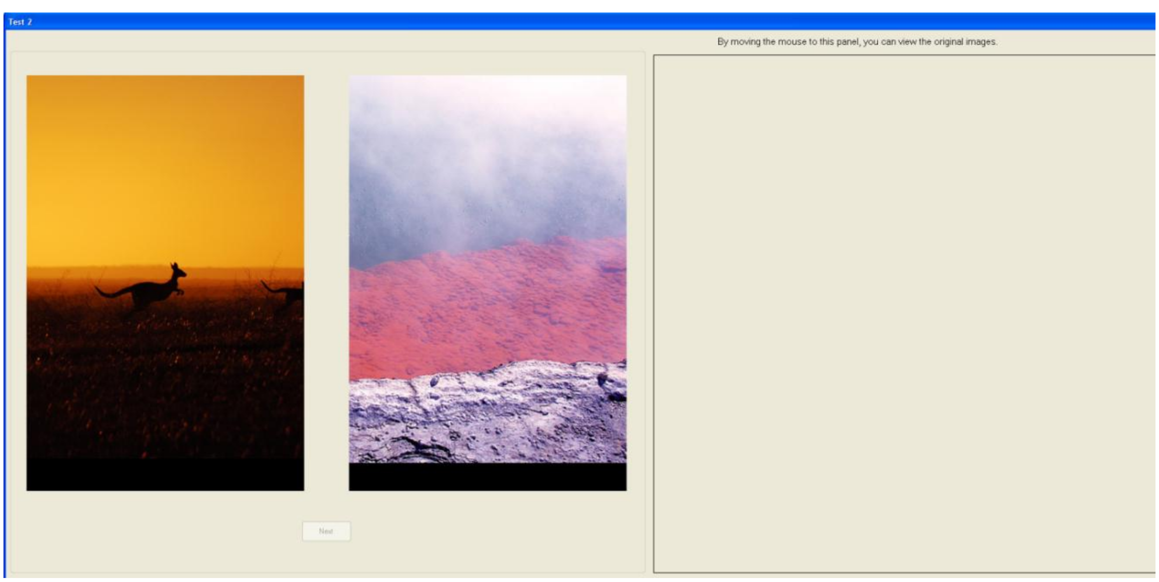
Figure 5-2 Flowchart of recognition task

Figure 5-2 shows the flowchart of the experimental task. In the first step, each subject viewed a block of 120 pictures. Each block comprised 60 "test pictures", which were used in later recognition tests and 60 "buffer item pictures" which were included to increase the memory load and bring the performance to middle range. The first and last 10 pictures of each block were all buffer items to eliminate the primacy and recency effect that might contaminate the subsequent memory test. The presentation sequence of the remaining 60 "test pictures" and 40 "buffer item pictures" were randomized.

The 60 "test pictures" were randomly grouped into 3 subsets of 20 pictures; 20 As, 20 Bs, and 20 Cs. In the two recognition tests, picture As were presented together with its similar counterpart A’s, resulting in 20 pairs of type AA’. Picture Bs were presented with the similar counterpart C’s of picture Cs, resulting in 20 pairs of type BC’.

After reviewing a block of 120 pictures, subjects performed an interpolated mathematical task for 5 minutes (http://pebl.sourceforge.net/). This task was used to bring the hit rate into the middle performance range.

After the interpolated task, subjects were given two recognition tests, each consisting of 20 questions of recognition (10 AA’ pairs & 10 BC’ pairs). In recognition test 1, subjects depended on their memory entirely and rated their confidence using a six-point scale. The points on the scale were 50%, 60%, 70%, 80%, 90% and 100%, with the leftmost point (50%) labeled “Just guessing” and the right most point (100%) “Absolutely confident”. In recognition test 2, subjects were allowed to access information in the world, and no confidence assessments were made. For any recognition questions, if they wished, subjects could access information in the world – the original pictures, by moving their mouse to a specific area.



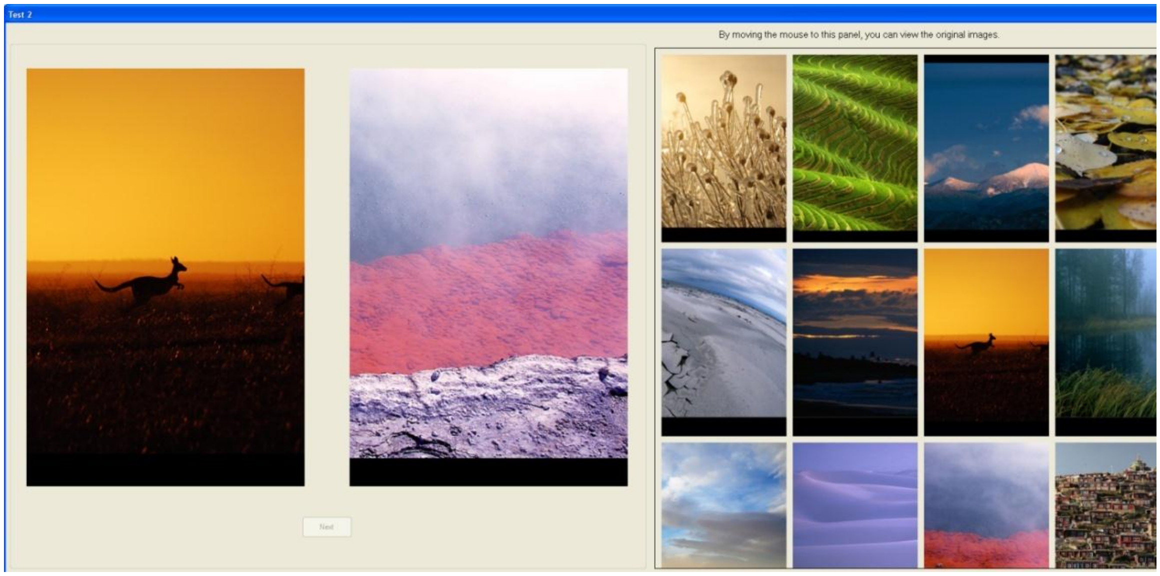


Figure 5-3 Mouse entering to uncover the original pictures

Participants repeated step 1 to step 4 three times. The three trials differed from each other in terms of the level of IAC provided in recognition test 2.

5.1.4 Experimental design

Independent variable The first independent variable was picture type (AA' vs. BC'), which as we noted changed the accuracy-confidence relationship; the second independent variable was IAC, which was manipulated by varying the delay time when accessing the original pictures (Gray et al., 2006). In the low IAC condition, the original pictures were shown immediately without any delay. In the medium IAC condition, there was 1 second delay and in the high IAC condition, 2 seconds delay.

Dependent variable In recognition test 1, accuracy and confidence judgment were measured as mediating variables, based on which overconfidence were calculated. In recognition test 2, number of information access attempts and recognition performance were measured.

Design The experimental design was a two by three within-subject design (Table 5-1). In every trial participants were given10 AA' pairs & 10 BC' pairs in each recognition test. The three trials were presented in a counterbalanced order such that 1/3 of the subjects received High IAC first, 1/3 received medium IAC first, and 1/3 received low IAC first.

Table 5-1 Experimental design for experiment 1

IAC / Picture type	No delay	Medium delay	High delay
AA'			
BC'			

5.1.1. Experimental apparatus and procedure

The experiment was coded in Visual Basic 2008 and used one 2-GHz personal computers. A 24 inch monitor was employed to make sure that the thumbnails of pictures were large enough for the subjects to see the differences. The activities of the mouse entering and leaving the area containing original picture in recognition test 2 were captured and time stamped by the Visual Basic program.

Upon entering the experimental room they were presented with the flowchart of the experiment and were given general information about the experiment. After that, they were given the following instruction:

"In the experiment you will view some pictures and later on recognize them in two tests. It consists of four steps. At step 1, you will view 120 pictures, with each picture showing for 2 seconds in the computer screen. Please pay attention to the pictures because later on your recognition memory of the pictures will be tested.

At step 2, you will perform a mathematical task for 5 minutes. Please do it as accurately as possible and as fast as possible.

At step 3, you will perform the first recognition test. In the test, you will be presented with 20 pairs of pictures. Each pair includes one picture you have seen before (target picture) and one distractor. The distracter is in one of the two forms (1) it is similar to the picture target picture in the pair (2) it is not similar to the target picture in the pair, but similar to another picture you have seen before. Your goal here is to select the picture you have seen before and indicate the confidence of your answer. If you are absolutely certain of your answer, select 100%. If you are just guessing, select 50%. The *higher the number, the more confident you are. In this test, we would like you to do it as accurately as possible.*

At step 4, you will perform the second recognition test. You will be presented with another 20 pairs of pictures and similarly your goal is to select the one you have seen before. However, in this step, you are provided with a memory aid located at the right panel of the software. In case you cannot remember a picture clearly, you can access the 120 pictures you have seen in step 1 by moving your mouse to the panel at the right side. The panel is covered up initially and once you move your mouse to the panel the 120 pictures will be shown (Figure 5-3). Moreover, once the mouse leaves the panel, all the pictures will be covered up again. In this test, we would like you to try your best to achieve at least 90% accuracy. At the same time, please do it as accurately and as fast as possible.

In the experiment, you will view 3 blocks of 120 pictures and therefore perform step 1 to step 4 three times. The procedures are exactly the same except you will experience different delay time in the second recognition test. In the low delay block, the original pictures will show immediately after you move your mouse to the panel. In the medium delay block, there will be 1 second delay before the pictures appear and in the high delay block, 2 seconds delay.

After each block, you will be given a 5-minute break."

After the orientation, the subjects completed a practice session with 30 pictures. The experimenter was present in the practice session to make sure the subjects were clear about the task. The experimenter left the experimental room when the subjects started the actual tasks. No feedback on performance was given to the subjects after each block.

It was observed in the pilot study that participants' recognition test 1 was relatively high. Therefore, to avoid the possibility that participants did not access information in the world in recognition test 2, a sentence stressing the performance requirement in recognition test 2 was added in the task briefing ("*In this test, we would like you to try your best to achieve at least 90% accuracy.")*.

5.2 Results

5.2.1 Analysis on the accuracy-confidence inversion

In recognition test 1, memory accuracy and confidence judgment were measured. Confidence was calculated by averaging the confidence judgment of question i (i =1 to N)

$$\mathrm{Confidence} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{Confidence}_{\mathrm{i}}$$

Overconfidence occurs when one's subjective belief of his/her judgment is greater than the objective accuracy and underconfidence occurs in the reverse condition (Koriat, Lichtenstein, & Fischhoff, 1980). Over/underconfidence was calculated as the signed difference between consolidated confidence and accuracy. A positive sign indicates overconfidence and a negative sign underconfidence.

$$\mathrm{Over/underconfidence} = \mathrm{Confidence} - \mathrm{Accuracy}$$

A two-way ANOVA was conducted to test the effect of picture type and IAC. Table 5-2 showed the effect of picture type on accuracy, confidence and overconfidence. There is a significant effect of picture type on hit rate ($p < .05$) and overconfidence ($p < .05$).

However, there was no significant difference between the two picture types in confidence judgments. According to the study of Tulving (1981), the pattern of accuracy-confidence inversion was expected. Thus the results were only partially consistent with the expectations.

Table 5-2 Mean value, F value and significance level of accuracy, confidence and overconfidence for picture type AA' and BC'

DV	Type AA'	Type BC'	F value	Sig.
Accuracy	8.074	7.556	F (2,35) = 7.244	0.011
Confidence	78.86	78.94	F (2,35) = 0.001	0.982
Overconfidence	-1.758	3.38	F (2,35) = 7.501	0.010

A detailed examination of the data suggested that the trial sequence seemed to affect accuracy and confidence judgment. Therefore, a two-way ANOVA was conducted treating IAC and trial sequence as the independent variables. Table 5-3 showed the mean value, F value and significance level for each dependent variable. The results revealed a significant interaction effect between picture type and trial sequence on confidence judgment and overconfidence.

Table 5-3 Mean value, F value and significance level of accuracy, confidence and overconfidence for picture type and trial sequence

IV	DV	F value	Sig.	Partial Eta Squared
Picture type	Accuracy	F (2,35) = 7.244	.011	.171
	Confidence	F (2,35) = 0.002	.967	.000
	Overconfidence	F (2,35) = 7.501	.010	.176
Trial sequence	Accuracy	F (2,35) = 2.394	.099	.064
	Confidence	F (2,35) = 8.661	.000	.198
	Overconfidence	F (2,35) = 0.302	.740	.009
Interaction	Accuracy	F (2,35) = 1.597	.210	.044
	Confidence	F (2,35) = 5.353	.007	.133
	Overconfidence	F (2,35) = 4.750	.012	.120

To investigate the significant interaction effects, a follow-up simple effect analysis was conducted. Table 5-4 shows the mean value of accuracy, confidence and overconfidence. These data are pooled over IAC subgroups.

Table 5-4 Mean value of accuracy, confidence and overconfidence for picture type AA' and BC' at each level of trial sequence

Trial sequence	Picture type	Accuracy (%)	Confidence (%)	Overconfidence (%)
1st trial	AA'	79.72	75.04	-4.68
	BC'	71.39	78.79	7.40
2nd trial	AA'	82.22	79.94	-2.28
	BC'	75.56	77.64	2.08
3rd trial	AA'	80.28	81.93	1.65
	BC'	79.72	80.39	0.66

The results suggested that the accuracy-confidence inversion pattern was prominent in the first trial, diminished in the second trial and was entirely eliminated by the third trial. The detailed analyses of these data are shown below.

The accuracy levels were higher in picture type AA' than BC' in all three trials, with decreasing differences. The differences were significant in the first trial (difference = 8.33, $p < .01$) and the second trial (difference = 6.67, $p < .05$) but not the third trial (difference = 0.56, $p = .889$).

The confidence showed a significant difference in the first trial (difference = -3.74, $p < .05$), with lower confidence in picture type AA'. In the second and third trials there was no significant difference in confidence between the two picture types (2nd trial: $p = .132$, 3rd trial: $p= .227$).

The results on overconfidence showed that the subjects were generally more overconfident in picture type BC' than AA'. In the first trial, the difference was significant (difference = -12.08, $p < .01$). As subjects proceeded to the second trial, the difference decreased (difference = -4.36, $p = .107$). In the third trial, there was essentially no difference in overconfidence (difference = 0.99, $p = .795$).

5.2.2 Analysis on the effect of IAC, confidence judgment and overconfidence

Due to significant effect of trial sequence on the accuracy-inversion pattern, it was not meaningful to analyze the data following the original 2 by 3 within subject design. Therefore, a one-way ANOVA was performed to analyze the effect of IAC solely. Table 5-5 shows the mean value of information access attempts and recognition test 2 performance results.

Table 5-5 Mean value of information access attempts (IAAs) and recognition performance

IAC	Number of IAAs	Recognition test 2 performance (out of 10)
Low	5.92	8.83
Medium	5.83	9.04
High	5.00	8.87

The analysis revealed a significant effect of IAC on the number of information access attempts ($F(2,70) = 4.609$, $p < .05$). That is, as the IAC increased from low and medium to high condition, the number of information access attempts decreased significantly. However, there was no effect of IAC on recognition test 2 performance results ($F(2,70) = .880$, $p = .419$). Pairwise comparison between the three levels of IAC showed that there was no difference in information access attempts between low and medium IAC (Low IAC: 5.92 vs. Medium IAC: 5.83), and there was a significant difference between medium and high IAC (Medium IAC: 5.83 vs. High IAC: 5.00, $p < .05$).

As previously mentioned, the accuracy-confidence inversion was only significant in the first trial. Therefore it was feasible to analyze the effect of confidence judgment with the data from the first trial only. Note that the first block of the counterbalanced design included data of 12 subjects at each level of IAC. The data were pooled over the 3 levels of IAC for the analysis. Table 5-6 shows the mean value of confidence judgment and the two dependent variables in the analysis.

Table 5-6 Mean value of accuracy, confidence, overconfidence, information access attempts (IAAs) and performance with data of 1st trial in both recognition tests

Picture type	Hit rate (%)	Confidence (%)	Over-confidence (%)	Number of IAAs	Recognition test 2 performance (out of 10)
AA’	79.72	75.04	-4.68	6.19	9.11
BC’	71.39	78.79	7.40	5.14	8.33

A one-way analysis of variance (ANOVA) revealed a significant difference on number of information access attempts ($F(1,35) = 5.016$, $p < .05$) and on recognition test 2 performance ($F(1,35) = 9.050$, $p < .01$). As the confidence judgment and overconfidence increased for picture type BC’, i.e. subjects were more confident and more overconfident towards information in their head, they accessed information in the world less frequently and the performance worsen.

Within the first block, there were 12 subjects at each IAC level. Therefore, it was feasible to test the interaction effect between IAC and confidence judgment, despite a much smaller sample size. The experimental design was a 3 by 2 mixed design, with IAC the between subject variable and picture type the within subject variable. Table **5-7** shows the mean value of information access attempts at each cell. The two-way ANOVA revealed no significant interaction effect between IAC and picture type on information access attempts (F(2, 33) = 0.082, P=.921), nor interaction effect on performance (F(2, 33) = 0.209, P = 0.812)

Table 5-7 Mean value of information access attempts (IAAs) and performance at each level of IAC and picture type with data from 1st trial

IAC	Picture type	Number of IAAs	Recognition test 2 performance (out of 10)
Low	AA'	6.75	9.00
	BC'	5.83	8.25
Medium	AA'	5.75	9.08
	BC'	4.83	8.50
High	AA'	6.08	9.25
	BC'	4.75	8.25

5.3 Discussion

Experiment 1 aimed to examine the factors affecting people's information access behavior and performance. A number of previous studies demonstrated that the costs associated with various information access strategies are an important factor that determines one's interactive behavior and performance results (Gray & Fu, 2004; Gray et al., 2006). However, none of the studies took into consideration the varying levels of accuracy associated with these strategies, which were supported by the cost-benefit decision making theories (Payne et al., 1988; Wickens et al., 2004), to be another determinant factor for information access behaviors.

Six hypotheses were tested in the experiment. H1 tested the accuracy-confidence inversion pattern; H2 examined the effect of IAC information access attempts and recognition performance. H3 examined the effect of confidence judgment on information access attempts. H4 tested the effect of overconfidence on recognition performance. H5 and H6 examined the interaction effect between IAC and confidence judgment, between IAC and overconfidence on information access attempts and recognition performance, respectively.

The study utilized the accuracy-confidence inversion pattern reported in Tulving (1981) to “manipulate” the confidence judgment and overconfidence between two types of picture pairs. The IAC was manipulated by varying lockout time when accessing information in the world (Gray et al., 2006, Experiment). How the experimental results support the six hypotheses are discussed below.

H1: There is an inverse pattern of accuracy and confidence judgments between picture type AA’ and BC’. Picture type AA’ will produce higher memory accuracy but lower confidence, while picture type BC’ will produce lower memory accuracy but higher confidence.

The results supported the hypothesis of accuracy-confidence inversion. However, the inversion pattern was moderated by task experience. The inversion was shown in the first trial, which was diminished in the second trial and eliminated in the third trial (Table 5-4). The diminishment of the inversion pattern was not reported in Tulving (1981) or another similar study (Dobbins et al., 1998), as subjects in the two studies did not participate in viewing multiple trials.

In an attempt to explain factors underlying the inversion pattern, Tulving (1981) stated that it was possible that “Highly similar test items (picture type AA’) may induce subjects to engage in deeper or more elaborate processing of retrieval information” and hence result in better memory performance. Also, the author suggested that more empirical studies were needed to fully understand the inverse pattern.

Since the diminishment of accuracy-confidence inversion is not the focus of the present study, a thorough explanation is not within our scope. However, it should be noted that in picture type AA’, subjects could make a correct recognition only when they remember the studied picture A, whereas in picture type BC’, subjects could make a correct recognition as long as they remembered either of the two studied pictures (picture B or picture C). Therefore, it is possible that when subjects became more experienced, the memory traces of all pictures become deeper, supported by increasing memory accuracy from the 1st trial to the 3rd trial, the advantage of picture type AA’ (inducing deeper processing) in hit rate lessened and the advantage of picture type BC’ (double chance to make a correct recognition) outweighed.

An alternative explanation is that the confidence-accuracy calibration was learned across the 3 experimental trials. That is, training induces a “debiasing” (Koriat et al., 1980). Although no explicit feedback was given to participants in the experiment, the results suggested that participants had many opportunities to carry out self-induced debiasing techniques. Had the information access attempts all been made for pictures that participants could not recall, the performance results for both picture types would have been perfect (10 out of 10) (Table 5-5). Nevertheless, neither picture type achieved perfect recognition, which suggested that some access attempts were made on pictures the participants could recall. Such access attempts served as self-administered feedbacks and helped the debiasing process.

H2: As IAC increases, the number of information access attempts in the world will decrease and performance will decrease.

The results provide partial evidence for the effects of IAC on information access attempts. As IAC increased from 1s to 2s, the number of times the memory aid was opened decreased significantly. There was a shift in information access from the world to access from the head (Table 5-5).

It is important to note that the effects of IAC on information access attempts were not linear across the three levels. The low and medium IACs had equivalent number of information access attempts, whereas the high IAC had a decreasing number of information access attempts. The lack of linearity could have resulted because the extra 1s delay from low to medium IACs was below the subjects' threshold for strategic change.

Grey et al. (2006) used six levels of lockout time ranging from 0ms to 3200ms and reported a significant linear trend of across condition that accounted for 97% of the variance. A close examination of information access attempts at each level of lockout time from their data suggested that subjects did not reduce access attempts proportionally with an equivalent increase in lockout time (Table 5-8). Although the significance of difference between adjacent levels was not reported, it should be noted that an equal step change in lockout time did not result in a relatively equivalent change in information access attempts. The result could be due to the same reason – a change in lockout time may not exceed subjects' threshold for strategic change and hence will not cause large difference in information access attempts.

Table 5-8 Number of information access attempts (IAAs) at different levels of lockout duration (Source: Gray et al., 2006)

	Lockout duration in milliseconds					
	0	200	400	800	1600	3200
Number of IAAs	5.6	4.8	4.5	3.7	3.5	2.9

In addition, the effect of IAC on performance was not significant, which was inconsistent with past studies. It could have resulted from two resources. First, despite the effort trying to bring the accuracy (hit rate) to middle range, the performance level in recognition test 1 was fairly high. The ceiling effect could have suppressed the effect of IAC. Second, the effect of IAC on information access attempts was only partially supported, which could have suppressed its influence on performance. A larger difference in IAC would be expected to lead to more prominent results.

H3: As confidence judgment increases, the number of information access attempts in the world will decrease.

Results from study 2 confirmed H3. When the overconfidence judgment increased from picture type AA' to picture type BC', despite a reversed pattern in memory accuracy, the number of information access attempts decreased significantly (Table 5-6). It is important to note that the information access attempts were determined by the 'metacognition' of memory accuracy instead of the actual performance.

H4: As overconfidence increases, recognition performance will decrease.

Study 2 results provided evidence of H4 that increasing overconfidence leads to reduced performance (Table 5-6). If memory accuracy and confidence judgment are well calibrated, any change in information access behaviors caused by change in confidence can be considered as a rationale adaptation. However, in reality, miscalibration between accuracy and confidence is usually the case. In the condition of overconfidence, humans depended more on their memory than they should have, which resulted in retrieval of erroneous information and harm the recognition performance. On the other hand, in the condition of underconfidence, humans trusted their memory less than they should have, resulting in more information access attempts in the world than needed.

H5: There is an interaction between IAC and confidence judgment on information access attempts. As IAC increases, the effect of confidence judgment on information access attempts will be more prominent.

The result of experiment 1 did not provide evidence for the interaction effect. The lack of significance could have resulted from the small sample size, given that the influence of the confidence inversion was only observed in the first trial, which reduced the number of subjects per IAC level. However, it should be noted that there was a trend of increasing difference in information access attempts, though not significant, as IAC increased from low and medium (Difference = 0.92) to high (Difference =1.23) (Table 5-7). It is likely that with a larger sample size and/or a larger difference in IAC, an interaction effect would have been found. To fully understand the interaction effect, more empirical studies are needed.

H6: There is an interaction between IAC and overconfidence on recognition performance. As IAC increases, the effect of overconfidence will be more prominent.

The result of study 2 did not provide evidence for the interaction effect between IAC and overconfidence on recognition performance. Similar to the explanation of H5, the lack of significance could have resulted from the small sample size at varying IAC levels. More empirical studies are needed to understand the interaction effect.

In conclusion, Study 2 provided evidence that besides information access cost, the confidence judgment is another important determinant for information retrieval behaviors. As confidence judgment increased, there were fewer information access attempts. Study 2 failed to provide evidence for the interaction effects between IAC and confidence judgment, and between IAC and overconfidence, partially due to the small sample size. The takeaway points of Study 2 are summarized below.

- Information access cost affects negatively the information access attempts to knowledge in the world (KIW)
- Confidence judgment affects negatively the information access attempts to knowledge in the world (KIW)
- Overconfidence bias has a negative impact on recognition performance.

6. Study 3: Effect of information access cost and accountability on medical residents' information retrieval strategy and performance at pre-handover

Study 2 investigated the factors that influence people's behavior in accessing information and reported the significant effects of IAC, confidence judgment and overconfidence on information access attempts and recognition performance. However, due to the unexpected significant effect of trial sequence and the resulting small sample size, the interaction effects between IAC and confidence judgment, between IAC and overconfidence were not fully examined.

This chapter presents an experimental study simulating medical residents' activities in the pre-handover phase. The aims of the experiment were two folded. First, it aimed to apply the same research model to examine medical residents' information retrieval strategies. Second, it aimed to investigate again the interaction effects between IAC and overconfidence bias.

Among the factors which have been shown to influence memory accuracy, confidence and overconfidence (see section 2.5), accountability was chosen to "manipulate" accuracy, confidence judgment and overconfidence (Figure 6-1), largely due to its prospective practical implication - it could be introduced as an intervention tool to improve patient safety in hospitals. Several studies provided consistent evidence that increasing accountability improves accuracy and reduces overconfidence (Ronis & Yates, 1987; Sniezek et al., 1990; Tetlock & Kim, 1987), though its effect on confidence judgment was inconclusive. Same as Experiment 1, IAC was the other independent variable, which was manipulated by imposing a 5-meter distance in the high IAC level.

There were two dependent variables - number of information access attempts and pre-handover performance. Study 3 aimed to test the following four hypotheses.

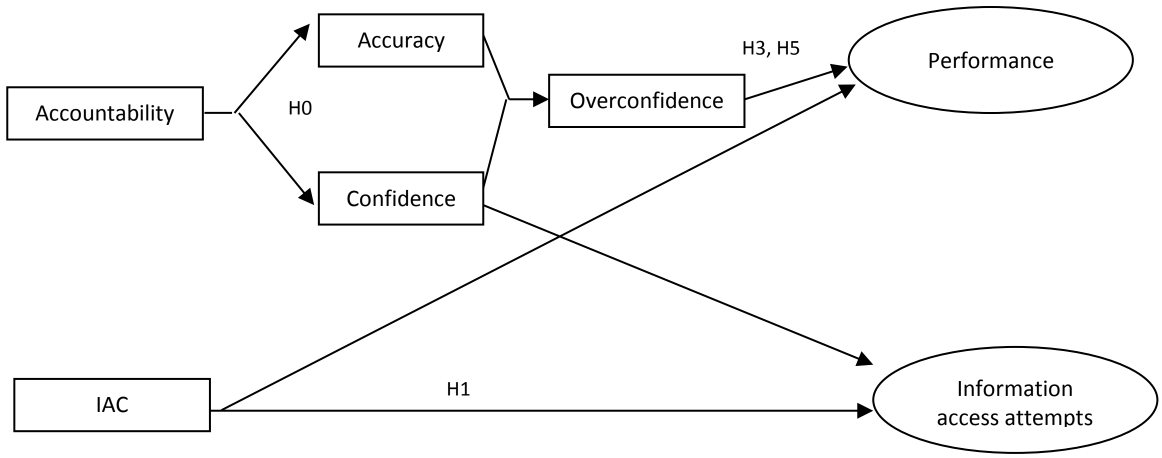


Figure 6-1 Research model between accountability, IAC, confidence, information access attempts and performance

H1: As accountability increases, memory accuracy will increase and overconfidence will decrease.

H2: As IAC increases, the number of information access attempts will decrease and the performance will decrease.

H3: As overconfidence decreases, the performance will increase.

H4: There is an interaction between IAC and overconfidence. As IAC increases, the effect of overconfidence will be more prominent.

6.1 Method

6.1.1 Stimuli

Based on real patient cases, A SME panel consisting of one consultant, one registrar and one house officer developed 12 mock up patient cases covering a wide range of subspecialties, including gastroenterology, neurology, respiratory, and cardiology. Each case consisted of patient photos, patient medical record and investigative test results (Figure **6-2**). Furthermore, for each patient case, the SME panel developed 10 recognition questions and 10 recall questions that were critical for the understanding of the patient. An example recognition questions is *"Was the patient given Plavix on admission?" and an example recall question is "Besides chest pain, what is the other symptom the patient was present with on admission?"* The 20 questions were used in the two recognition/recall tests.

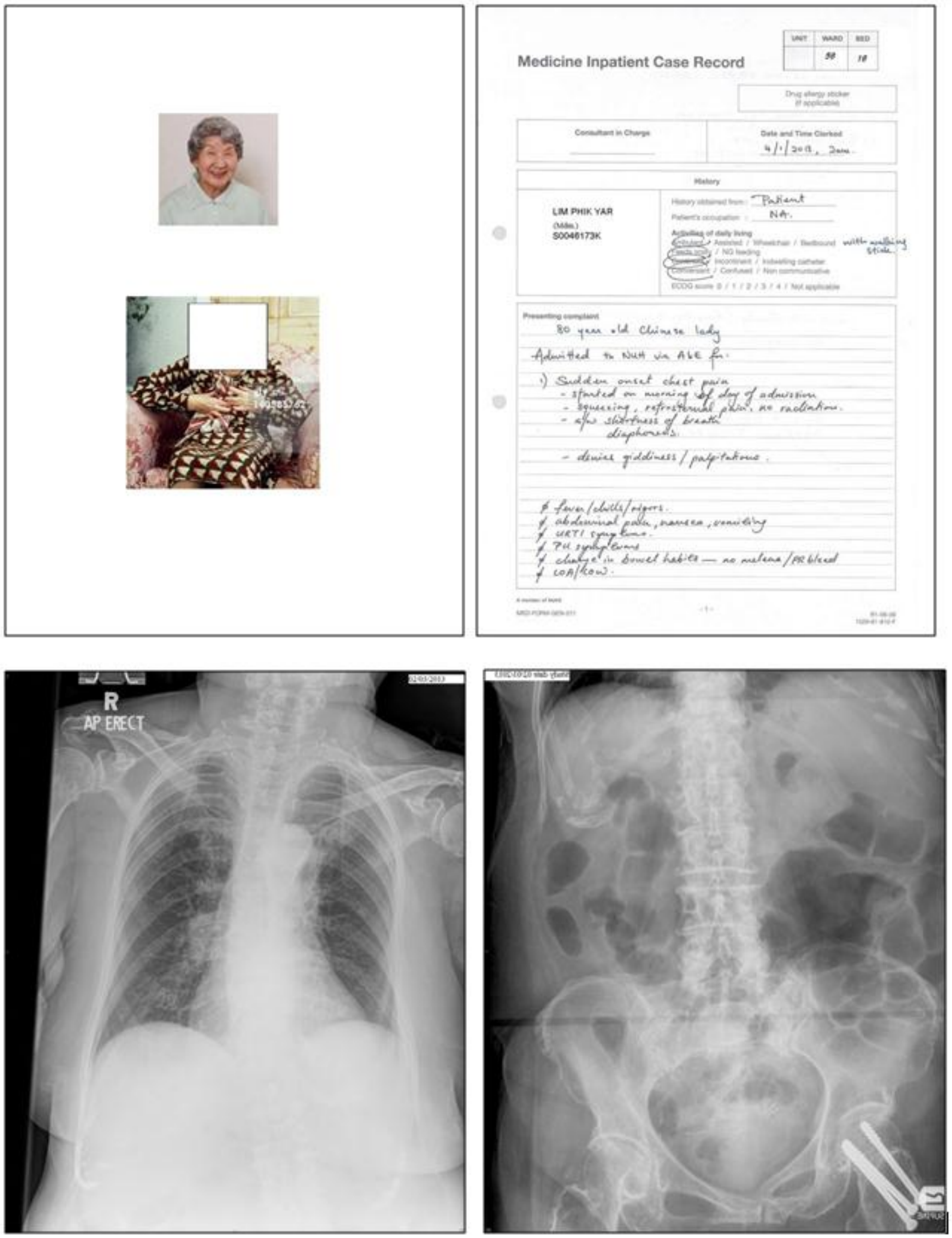


Figure 6-2 A sample patient case comprising patient photos (80 years old Chinese lady with sudden onset of chest pain), medical record and investigative tests

The 12 patient cases were categorized into two groups according to their information complexity. Complexity ratings were assessed on a scale of 1 (simple) to 5 (difficult) by the registrar and the house officer. The complex cases had an average of 4.2 (SD = 0.6) and the simple case had an average of 3.3 (SD = 0.7) on the difficulty scale. The 12 cases were then divided into two LOT A and LOT B based on two criteria. First, each lot consisted of 3 complex cases and 3 simple cases. Second, no two cases from a lot were from the same subspecialty.

The 12 mock up patient cases were used as experimental materials for Study 3 and Study 4 (see Chapter 7). For Study 3, 4 patient cases from each LOT were selected as experimental materials. The 4 four patient cases from each lot consisted of 2 complex cases and 2 simple cases.

6.1.2 Participants

Thirty-two residents including medical officers and house officers from Changi General Hospital (CGH) and Alexandra Hospital (AH) participated in the study. The National Healthcare Group (NHG) Domain Specific Review Board (DSRB) and Changi General Hospital (CGH) Internal Review Board (IRB) approved the study and all participants gave informed consent. The results of 2 participants were excluded from the analysis. One of them chosen confidence 0 throughout the entire study and the other answered “not sure” for 90% of the recall questions. The remaining 30 subjects had an average age of 28.1 years (SD = 3.1 years) and an average experience of 3.6 years (SD = 2.5 years). Twenty six of the 30 participants were from General Medicine department, 3 from General Surgery department and 1 from Anesthesia department. Each participant was paid 50 Singapore dollars (39.4 USD) for their participation.

6.1.3 Experimental task

Figure 6-3**Figure 5-2** shows the flowchart of the study. Each participant was randomly assigned to view the patient cases from LOT A or LOT B.

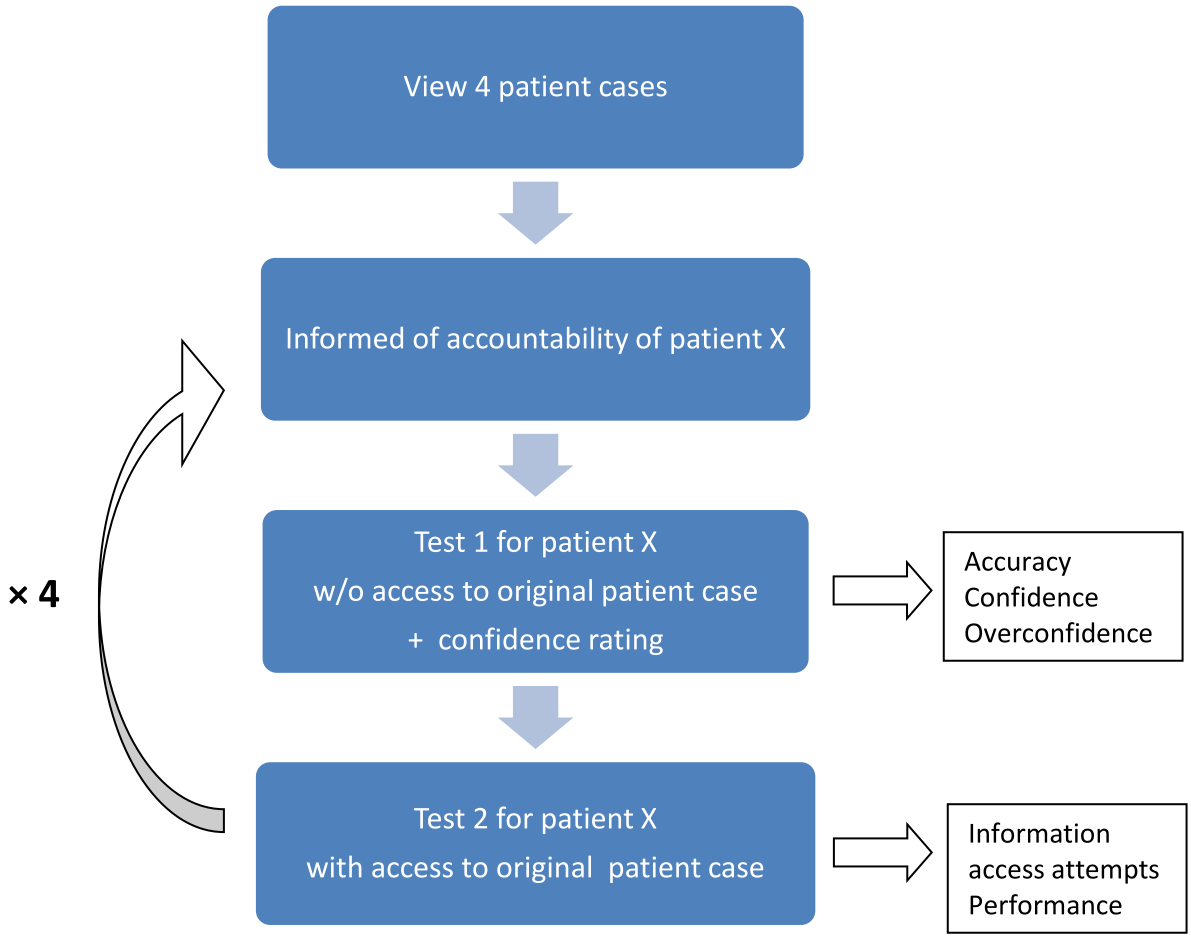


Figure 6-3 Flowchart of study 3

After viewing the 4 patient cases, accountability of a particular patient (patient X) was announced prior to test 1 and test 2 (Figure 6-4). Adopting from the study of Tetlock and Kim (1987), in the low accountability condition, subjects were assured that their performance would be completely confidential, whereas in the high accountability condition they learned that their test results would be assessed and revealed at the end of the study.

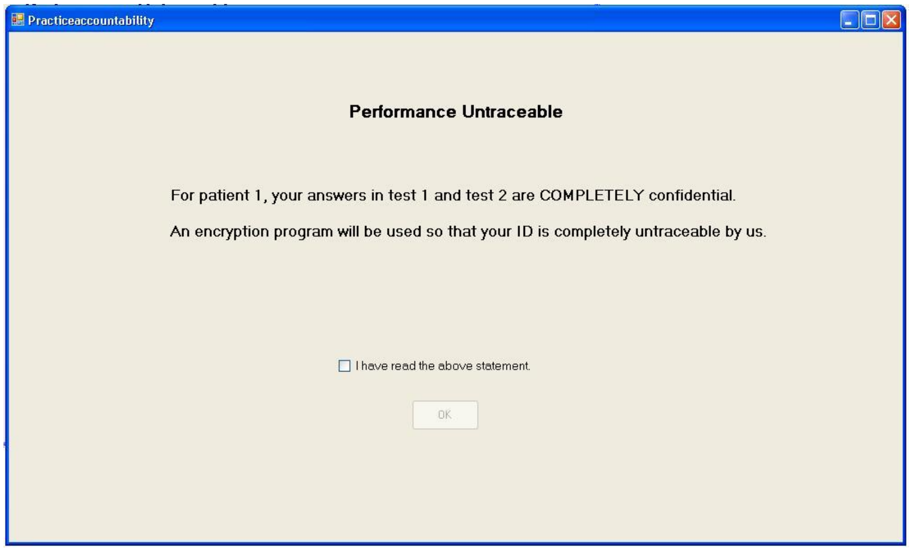


Figure 6-4 Low accountability for patient 1

After that, subjects were given the two tests. Each test consisted of 5 recognition questions and 5 recall questions. In Test 1 participants depended on their memory entirely and provided confidence estimates. After each recognition question, subjects rated their confidence in the answer using a six-point scale. The points on the scale were 50%, 60%, 70%, 80%, 90% and 100%, with the leftmost point (50%) labeled "Just guessing" and the right most point (100%) "Absolutely confident". After each recall question, subjects rated their confidence using an eleven-point scale, ranging from 0% to 100%, with the leftmost point (0%) labeled "Just guessing" and the right most point (100%) "Absolutely confident".

In Test 2, participants were allowed to access KIW. For any test question, if they wished, subjects could access the corresponding original case file, by clicking a button either at the PC where the tests were undertaken or a PC located 5-meter away. The activities of the mouse clicking were captured and time stamped. Participants repeated step 2 to step 4 for the 4 patients.

A pilot study with 2 junior doctors was carried out to determine the values of two variables. First, in order to bring the hit rate into the middle performance range, the viewing time was manipulated and the results suggested that 150 seconds for each patient was appropriate. Second, in order to ensure that the average difficulties of the questions in Test 1 and Test 2 were balanced, stratified randomization instead of simple randomization was used. The difficulty score of each recognition question and recall question was obtained (based on memory accuracy results from Test 1). The difficulty score was used as the ***prior*** to stratify the questions into hard, moderate and easy questions. The assignment of the 20 questions into Test 1 and Test 2 were then randomized within each stratum. Moreover, the difficulty score was constantly updated after a participant completed the experiment.

6.1.4 Experimental design

Independent variable There were two independent variables: accountability and IAC. Accountability was the extent to which participants needed to justify their performance. It was varied across 2 levels. In the low level, subjects were assured prior to test 1 and test 2 that their performance would be completely confidential. Even the experimenter, it was emphasized, would not know how they had responded. In the high level, subjects were informed that their test results would be tracked down and the experimenter would discuss their performance with them at the end of the study. The second independent variable, IAC from the world, was the level of physical effort when accessing original patient cases. In the low level, the original patient cases were stored at the same PC that was used to perform the tests, whereas in the high level, the patient cases were stored at a PC located 5-meter away.

Dependent variable Several variables were measured in the two tests. In test 1, accuracy and confidence judgment were measured as the mediating variables, based on which overconfidence was calculated. In test 2, number of information access attempts and pre-handover performance were measured.

Design The experiment followed a 2 by 2 within-subject design. A 4 by 4 reduced Latin square was employed so that each patient case had equal possibility (25%) to be assigned to a certain condition. The studying sequence and the testing sequence of the 4 patients were randomized and counterbalanced.

Conditions		Patient cases				
Accountability	IAC					
No	Low	Patient 1	Patient 2	Patient 3	Patient 4	
No	High	Patient 2	Patient 1	Patient 4	Patient 3	
Yes	No	Patient 3	Patient 4	Patient 1	Patient 2	
Yes	High	Patient 4	Patient 3	Patient 2	Patient 1	
		Subject 1	Subject 2	Subject 3	Subject 4	…

6.1.5 Experimental apparatus and procedure

The experiment was coded in Visual Basic 2008 and used two 2.1-GHz personal computers with two 24-inch color monitors. The activities of the mouse clicking for original patient cases in Test 2 were captured by the Visual Basic program.

Upon entering the experimental room participants were presented with the flowchart (Figure 6-3) and were given general information about the study. After that, they were given the following instruction:

"In the experiment you will view 4 patients similar to the ones you see every day and later on prepare the handover of the 4 patients by answering some questions. The experiment consists of four steps. At step 1, you will view 4 patient cases consisting of patient photos, patient medical record and all the investigation test results. You will be given 150 seconds to view each patient. Please pay attention because later on you need to answer some questions about each patient.

After reviewing the four patients, you will come to the test session. Prior to the tests of each patient, you will know for that particular patient, whether your performance in the following two tests will be traceable. In the performance untraceable condition, your performance for the particular patient was completely confidential. An encryption program will be used so there is no way for us to track your performance. In the performance traceable condition, we are able to track your performance and we will review your performance together with you at the end of the study.

At step 3, you will perform the first test. In the test, you need to answer 10 patient related questions. There are two types of questions, open-ended questions and yes or no questions. A sample open-ended question is "Besides chest pain, what is the other symptom the patient is present with?" and a sample yes or no question is "Was the patient given Plavix on admission?" After answering each question, please indicate your confidence. For an open ended question, if you are absolutely certain of your answer, *select 100% and if you are just guessing, select 10%. For a yes or no question, if you absolutely certain of your answer, select 100% and if you are just guessing, select 50%. The higher the number, the more confident you are. In this test, we would like you to do it as accurately as possible.*

At step 4, you will perform the second test. You will be presented with another 10 patient related questions. However, in this step, you are allowed to access the original patient cases if needed. For a particular patient, the original patient case was either at hand or faraway. In the patient case is at hand, the original patient file is located at this computer (pointing to the same computer where participants will perform the tests). What you need to do is to press the "show case file" button and the original case will show up. If the patient case is faraway, the original patient case is located at that computer (pointing to the computer located 5-meter away. What you need to do is to physically walk there and press the corresponding button to view the original patient file. In this test, we would like you to do it as accurately and as fast as possible."

After the orientation, the subjects completed a practice trial. The experimenter was present in the practice session to make sure the subjects were clear about the task. The experimenter left the experimental room afterwards. No feedback on performance was given to the subjects after each block.

6.2 Results

Memory accuracy and confidence judgment were measured. Confidence and overconfidence were calculated by the same formula

$$\text{Confidence} = \frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} \text{Confidence}_{\text{i}}$$

$$\text{Overconfidence} = \text{Confidence} - \text{Accuracy}$$

6.2.1 Analysis on the effect of accountability on accuracy and overconfidence

A two-way ANOVA was conducted to test the effect of IAC and accountability. Table 6-1 shows effect of accountability on accuracy and overconfidence. Accountability significantly affected accuracy ($F(1,29) = 7.695$, $p < .01$, $\eta_p^2 = .210$) and overconfidence ($F(1,29) = 4.395$, $p < .05$, $\eta_p^2 = .132$) but not confidence judgment ($F(1,29) = .249$, $p = .622$) When accountability was high, subjects showed higher memory accuracy and smaller overconfidence.

Table 6-1 Mean value of accuracy, confidence, overconfidence (OC), information access attempts (IAAs) and Test 2 performance

Accountability	IAC	Accuracy (%)	Confidence (%)	OC (%)	Number of IAAs	Test 2 Performance
Low	Low	54.7	69.2	14.5	6.3	78.7
	High	55.7	68.9	13.3	1.5	62.3
High	Low	59.3	68.2	8.9	6.0	78.3
	High	63.0	71.1	8.1	2.1	69.0

6.2.2 Analysis on the effect of IAC and overconfidence

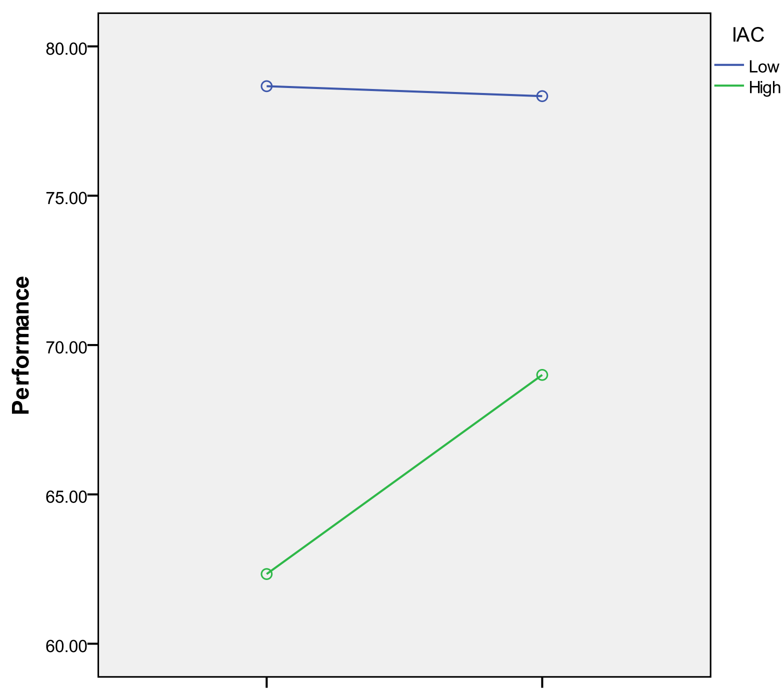
The two-way ANOVA also revealed significant effects of IAC on number of information access attempts ($F(1,27)= 97.79$, $P< .001$, $\eta_p^2=.784$) and on pre-handover performance (Test 2 performance) ($F(1,29)= 20.23$, $P< .001$, $\eta_p^2=.411$) (Table 6-1). When

IAC increased, participants accessed original patient cases less and pre-handover performance (Test 2 performance) decreased.

There was no main effect of overconfidence on performance (F(1,29)= 2.706, P= .111) nor interaction between IAC and overconfidence on performance (F(1,29)= 2.391, p= .133). However, the results suggested a trend that high accountability led to high performance when IAC was high (

Figure **6-5**). Therefore, a simple effect analysis was conducted to test the effect of overconfidence at different levels of IAC.

The simple effect analysis results revealed a significant difference in performance when IAC was high (Difference = 6.667, p< .05, 95%CI: .286, 13.05). When original patient cases were not easily accessible, smaller overconfidence resulted in better performance.



Large overconfidence Small overconfidence

Figure 6-5 Interaction effect between IAC and overconfidence

6.3 Discussion

Four hypotheses were formed and test in Study 3. H1 tested the effect of accountability on accuracy and overconfidence. H2 examined the influence of IAC on information access attempts and task performance. H3 tested the effect of accountability (and hence overconfidence) on performance. H5 tested the interaction between accountability (and hence overconfidence) and IAC on task performance. Below we discuss how the experimental results support the four hypotheses.

H1: As accountability increases, memory accuracy will increase and overconfidence will decrease

The results from Study 3 supported the hypothesis. As the accountability condition increased from performance untraceable to performance traceable, participants had higher hit rate and smaller overconfidence (Table 6-1), which was consistent with past results (Ronis & Yates, 1987; Sniezek et al., 1990; Tetlock & Kim, 1987). Furthermore, we noted that there was no difference in confidence judgment when accountability varies. The increase in overconfidence was defined by a loss in accuracy in low accountability condition without a loss in confidence, reflecting a faulty calibration of meta-cognition (Wickens et al., 2012).

H2: As IAC increases, the number of information access attempts will decrease and the performance will decrease.

Consistent with Experiment 1, the results provided evidence for the effects of IAC on information access attempts and performance. By imposing a 5-meter distance in the IAC high condition, the number of access to the original patient cases decreased significantly. There was a clear shift from access from the world to access from the head (Table 6-1). Further, the shift in information access strategy resulted in compromised performance.

Past studies have used various methods to manipulate IAC, including eye movement vs. mouse movement, varying levels of Fitts' index and varying levels of lookout time (Ballard et al., 1995; Gray & Fu, 2001, 2004; Gray et al., 2006; Morgan et al., 2009; Morgan et al., 2007). Study 3 showed that besides the above factors, physical distance is a viable way to inducing varying levels of IAC.

H3: As overconfidence decreases, the performance will increase.

Study 3 did not support the main effect of overconfidence on performance. The lack of significance could have resulted from the moderating effect of unnecessary information access attempts (See explanation of H4) .

H4: There is an interaction between IAC and overconfidence. As IAC increases, the effect of overconfidence will be more prominent.

Results from Study 3 provided partial evidence for the interaction effect (Table 6-1). Although the interaction between IAC and overconfidence on performance is non-significant, the simple effect analysis results showed a significant difference in performance when IAC was high (Difference = 6.667, $p < .05$).

A detailed examination of the results (Table 6-1) showed that when IAC was low, participants made far more access attempts to the original patient cases that they should had done. The average confidence in IAC low condition was 68.7%, which suggested that

3.13 information access attempts were completely necessary in order to achieve perfect performance. However, the participants had average access attempts of 6.15. The extra attempts moderated the effect of overconfidence and resulted in an insignificant difference in performance. On the contrary, when IAC was high, the incurred cost prohibited the unnecessary information access attempts and the difference in performance emerged.

In summary, consistent with Study 2, Study 3 showed that the negative effect of IAC on information access attempts and task performance and provided partial evidence for the interaction effect between IAC and overconfidence. However, it should be noted that the mock up patients in Study 3 were simulated through patient photos, patient case files and diagnostic test results. The simulation was limited in its fidelity, since the interaction with a patient in real settings usually takes place through more modalities, such as through auditory and tactile channels, and usually lasts for a much longer time. The differences in interaction experiences may affect the pre-handover and handover communication performance. The takeaway points of Study 3 are summarized below.

- Information access cost affects negatively the information access attempts to knowledge in the world (KIW) and recognition/recall performance.
- When information access cost is high, overconfidence bias has a negative impact on recognition performance.

7. Study 4: Effect of four media affordances on medical residents' handover communication

This chapter presents an experimental study simulating medical residents' handover communication. The aim of the study was to test the influence of four affordances on handover receivers' memory retention of patient information and comprehension of patient case, as well as the interaction effects between interactivity, reviewability, visual access to object of reference and task complexity. The hypotheses are listed as follows:

H1a: Communication medium supporting higher reviewability leads to better memory retention and case comprehension.

H1b: When case complexity is high, reviewability will more positively enhance memory retention and case comprehension

H2a: Communication medium supporting higher interactivity leads to better memory retention and case comprehension.

H2b: When case complexity is high, interactivity will more positively enhance memory retention and case comprehension.

H3a: Communication medium supporting access to object of reference leads to memory retention and case comprehension.

H3b: When case complexity is high, access to object of reference will more positively enhance memory retention and case comprehension.

H4: Communication medium supporting visibility does not lead to better memory retention and case comprehension.

7.1 Methods

7.1.1 Stimuli

The same pool of 12 mock up patient cases was used in the experiment. For each patient case, the SME panel developed 6 memory retention questions and 2 case comprehension questions. Knowing the answers of the memory retention questions were vital for completing the handed over task. A sample question was "What are the patient identifiers?" The case comprehension questions were developed to evaluate handover receivers' integrative understanding of a patient. A sample question was "What is the primary team's concern when they handed over the patient to you?"

7.1.2 Participants

The National Healthcare Group (NHG) Domain Specific Review Board (DSRB) and Changi General Hospital (CGH) Institutional Review Board (IRB) approved the study and all participants gave informed consent. Each participant was paid 50 Singapore dollars (39.4 USD) for their participation. Twenty-four of the 30 residents participated in Study 3 joined Study 4 (Age: 27.6 years (SD = 2.6 years), Experience: 3.4 years (SD = 2.5 years). The participants formed 12 pairs to role play the handover sender and the receiver. Each pair consisted of participant A who studied LOT A and participant B who studied LOT B in Study 3. The last 16 residents also participated in a subjective evaluation on the importance of the four affordances.

7.1.3 Experimental task

Figure 6-3 shows the flowchart of the study. In the first half of the study, one participant was randomly selected to be the handover sender, and the other the handover receiver. The handover sender was given 3 minutes to view one patient case from his/her own LOT. After viewing the patient, the sender handed over the patient to the receiver using a randomly assigned method (Table **7-1**). After the handover, the receiver was given 3 minutes to complete 6 memory retention and 2 case comprehension questions. The receiver also indicated subjectively his/her understanding of the patient and the to-do task. In the second half of the study, the two participants swapped their roles. The handover receiver in the first half of the study role played the sender, and vice versa.

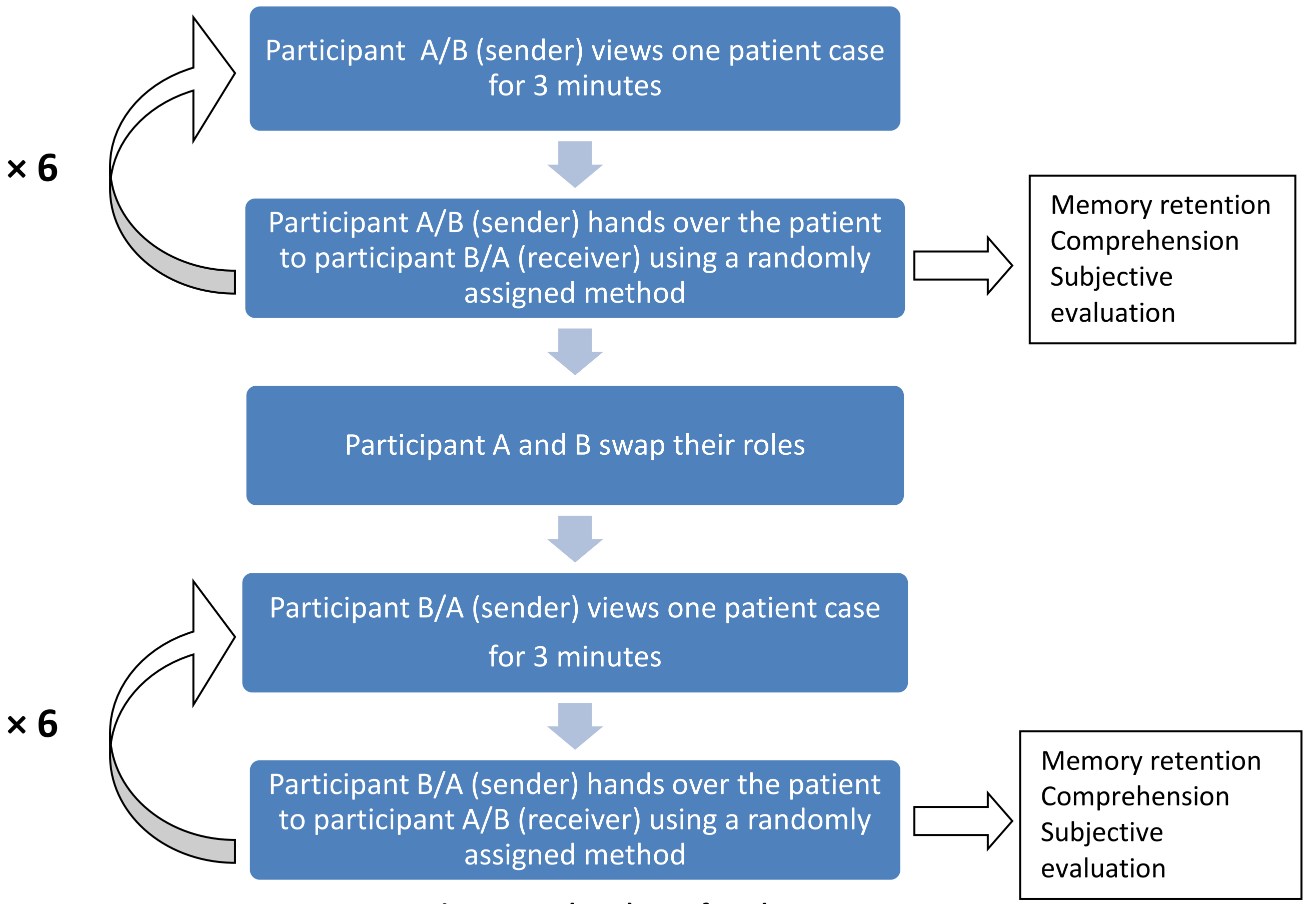


Figure 7-1 Flowchart of study 4

7.1.4 Experimental design

Independent variable There were five independent variables: reviewability, interactivity, access to object of reference, visibility and case complexity. The first four of these can be thought of as affordance attributes of different media, as shown below in Table 7-1. Reviewability was the extent to which handed over information persists over time and it was varied across the 3 levels. In the low level, handover receiver was not allowed to record down any patient information; in the medium level, the receiver was allowed to record down patient information on a blank piece of paper; in the high level, the handover sender prepared a handover note for the receiver prior to the communication and the receiver were allowed to write down addition information during handover communication if necessary.

Interactivity was the extent to which handover sender and receiver could communication in real time. In the low interactivity condition, the handover receiver was forbidden from asking questions and providing feedback. The receiver was asked to remain completely silent during the handover communication. In the high interactivity condition, the handover receiver was allowed to provide feedback and ask questions.

Access to object of reference was the extent to which handover sender and receiver could view patient related information simultaneously. In the “Yes” condition, handover receiver was provided with all patient related visual information, such as chest x-ray, ECG and MRI. In the “No” condition, handover receiver was not able to do so.

Visibility was the ability of seeing each other during handover communication. In the “Yes” condition, participants were allowed to see each other (but not the patient visual information) during handover communication and in the “No” condition they were not.

The last independent variable was case complexity. The complex cases had an average of 4.2 (SD = 0.6) and the simple case had an average of 3.3 (SD = 0.7) on the difficulty scale. In experiment 3, each of the four affordances was varied at both low and high levels of case complexity.

Manipulation of the four affordances resulted in 6 handover methods (Table 7-1).

TABLE 7-1 MANIPULATION OF AFFORDANCES AND THE CORRESPONDING HANDOVER METHODS

Affordance	**Interactivity** The extent to which handover sender and receiver can communication in real time.	**Reviewability** The extent to which handed over information retains over time.	**Shared workspace** Whether handover sender and receiver can view patient related information simultaneously.	**Visibility** Whether participants can see each other
Method A Receiver in Transit	Yes	Low	No	No
Method B Cell phone handover	Yes	Medium	No	No
Method C Cell phone plus handover note	Yes	High	No	No
Method D Audio Recording	No	Medium	No	No
Method E Face-to-face handover	Yes	Medium	No	Yes
Method F Distant consultancy	Yes	Medium	Yes	No

Cell Phone handover (Method B) is the most commonly used and accepted method in hospitals of Singapore currently (Yang, Park, Siah, Ang, & Donchin, 2014). Using this method, a handover sender and receiver communicate through cell phones and the receiver records vital patient information at his or her own discretion. Cell phone handover is considered as the baseline in the present study, and the other 5 methods are different from Cell Phone Handover by varying the condition for one of the four affordances.

Handover in Transit (Method A) and Written plus Verbal Handover (Method C) are different from Cell Phone Handover (Method B) in the affordance of reviewability. For Handover in Transit, handover receiver is not allowed to record any patient information. This method is used occasionally in hospitals in Singapore, when the handover receiver is in transit from place to place (Yang et al., 2014). In Written plus Verbal Handover, handover receiver is provided with written handover notes. Therefore he or she has full support of reviewability.

Handover using Audio Recording (Method D) is different from Cell Phone handover (Method B) in the affordance of interactivity. It simulates the condition of asynchronous communication. The handover receiver is not able to ask questions or clarify doubts when listening to the handover communication.

Compared to Cell Phone handover (Method B), Face-to-face handover (Method E) enables the handover sender and receiver to view each other, but not the patient or the patient case file.

Compared to Cell Phone handover (Method B), Handover using Distant Consultancy (Method F) enables the handover sender and receiver to view the patient and the patient case file, but not the facial expression or gesture of each other's.

Dependent variable Two rated performance variables were measured after each handover. Memory retention and case comprehension questions for each patient were marked independently by one registrar and one house officer. The inter-rater reliability was calculated using Spearman's correlation (r = 0.820), which showed good agreement between the two. Therefore the average value of the marks from the two evaluators was employed to measure memory retention and case comprehension. The subjective evaluation was measured by asking the handover receiver to indicate to what extend that "I completely understand the patient case and the to-do tasks" using a 7-point Likert scale (1 – not agree at all, 7 – completely agree)

Design Each participant handed over and received 3 complex patient cases and 3 simple cases using the 6 different methods (Table 7-2).

Table 7-2 Experimental design for each participant (Each participant followed either ○ or △)

	Complex	Simple
Method A Receiver in Transit	○	△
Method B Cell phone handover	△	○
Method C Cell phone plus handover note	○	△
Method D Audio recording	△	○
Method E Face-to-face handover	○	△
Method F Distant consultancy	△	○

A 6 by 6 reduced Latin square was employed so that each patient case had equal possibility to be assigned to a certain method. The presentation of handover methods was randomized.

Conditions	Patient cases					
Method A	Patient 1	Patient 2	Patient 6	Patient 3	Patient 5	Patient 4
Method B	Patient 2	Patient 3	Patient 1	Patient 4	Patient 6	Patient 5
Method C	Patient 3	Patient 4	Patient 2	Patient 5	Patient 1	Patient 6
Method D	Patient 4	Patient 5	Patient 3	Patient 6	Patient 2	Patient 1
Method E	Patient 5	Patient 6	Patient 4	Patient 1	Patient 3	Patient 2
Method F	Patient 6	Patient 1	Patient 5	Patient 2	Patient 4	Patient 3

Subject1 Subject2 Subject3 Subject4 Subject5
Subject6…

7.1.5 Experimental apparatus and procedure

The experiment was coded in Visual Basic 2008 and used two 2.1-GHz personal computers with two 24-inch color monitors (Figure 7-2).

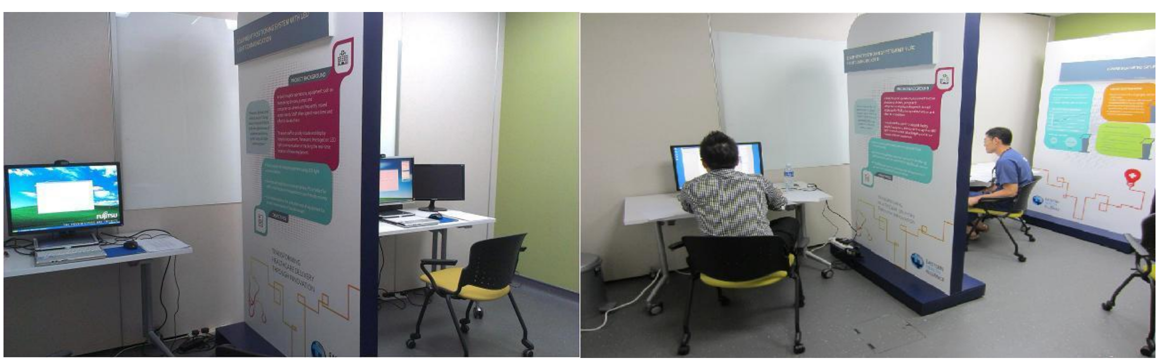


Figure 7-2 Experimental apparatus of Study 4

Upon entering the experimental room participants were presented with the flowchart (Figure 6-3) and were given general information about the study. After that, they were given the following instruction:

"In the experiment you will role play the handover sender and receiver to transfer and receive 6 patients using 6 different methods. In the first half of the experiment, one of you will be the sender and the other the receiver. As the handover sender, you will be given 3 minutes to study each patient. You can take notes during handover preparation. The patient case will be blanked out after 3 minutes. During the handover communication, the sender can always refer to the notes. Please note in one of the 6 methods, you will be asked to prepare a handover note for the receiver prior to the verbal handover.

As the handover receiver, during the communication you will be allowed to perform different actions according to each handover method. For example, in one method, you will be forbidden from providing feedback and asking questions. After receiving a patient, you will be given 3 minutes to answer a set of questions.

In the second half of the experiment, the two of you will swap your roles. The sender in the first half will become the receiver and vice versa."

The two participants were then given a table tabulating handover roles, and the sequence of patients and handover methods. The experimenter was present throughout the study to assist the two participants. Prior to each handover, the experimenter announced again the particular handover method to the two participants. At the end of the experiment, participants were asked to evaluate subjectively the importance of the four attributes of reviewability, interactivity, access to object of reference and visibility on a 7-point Likert scale (1-not important at all, 7 – extremely important).

7.2 Results

7.2.1 Effect of reviewability on handover quality

The effect of reviewability was tested by comparing method A and method C (Table 7-3). A two-way mixed ANOVA was conducted with reviewability as the within subject factor and case complexity as the between subject factor.

Table 7-3 Experimental design for testing reviewability

	Reviewability	Complex	Simple
Method A Receiver in Transit	Low	○	△
~~**Method B** Cell phone handover~~	~~Medium~~	~~△~~	~~○~~
Method C Cell phone plus handover note	High	○	△

As shown in Table 7-4, reviewability had significant benefits on memory retention ($F(1,22)= 9.453$, $p < .01$, $\eta_p^2= .301$) and subjective evaluation ($F(1,22) = 4.380$, $p <.05$, $\eta_p^2=.166$) but not on case comprehension ($F(1,22) = 0.023$, $p = 0.88$). Providing handover notes significantly improved handover receivers' memory retention and subjective evaluation. There were no interaction effects between reviewability and case complexity on any of the three measures (memory retention: $F (1,22) =0.979$, $p = .333$; case comprehension: $F (1,22) = .023$, $p =.88$; subjective evaluation; $F (1,22) =.294$, $p =.593$).

Table 7-4 Mean value of memory retention, case comprehension and subjective evaluation at each level of reviewability and case complexity

Reviewability	Case Complexity	Memory Retention (Out of 30)	Case comprehension (Out of 10)	Subjective evaluation
Low	Simple	21.5	7.1	5.3
	Complex	18.4	6.8	4.7

High	Simple	23.4	7.1	5.8
	Complex	22.2	7.0	5.4

7.2.2 Effect of interactivity on handover quality

The effect of interactivity was tested by comparing method B and method D. A two-way mixed ANOVA was conducted with interactivity as the within subject factor and case complexity as the between subject factor (Table 7-5).

Table 7-5 Experimental design for testing interactivity

	Interactivity	Complex	Simple
Method B Cell phone handover	High	△	○
Method D Audio recording	Low	△	○

The results shown in Table 7-6 revealed significant benefits of interactivity on memory retention ($F(1,22)= 6.522$, $p < .05$, $\eta_p^2= .229$) and case comprehension ($F(1,22) = 4.584$, $p < .05$, $\eta_p^2= .172$) and a non-significant effect on subjective evaluation ($F(1,22) = 2.112$, $p =.16$), but trending in the direction of higher evaluation for the more interactive method (Table 7-6). When receivers were allowed to provide feedbacks and ask questions, they demonstrated higher memory retention and case comprehension. No interaction effects with complexity were found (memory retention: $F(1,22) =1.562$, $p = .224$; case comprehension: $F (1,22) = .063$, $p =.803$; subjective evaluation: $F(1,22) =.528$, $p =.475$).

Table 7-6 Mean value of memory retention, case comprehension and subjective evaluation at each level of interactivity and case complexity

Interactivity	Case Complexity	Memory Retention (Out of 30)	Case comprehension (Out of 10)	Subjective evaluation
Low	Simple	18.3	5.7	5.1
	Complex	14.7	5.2	5.3

High	Simple	20.3	6.9	5.9
	Complex	20.6	6.8	5.6

7.2.3 Effect of visual access to object of reference on handover quality

The effect of interactivity was tested by comparing method B and method E. A two-way mixed ANOVA was conducted with method as the within subject factor and case complexity as the between subject factor (Table **7-7**).

Table 7-7 Experimental design for testing visual access to object

	Visual access to object of reference	**Complex**	**Simple**
Method B Cell phone handover	No	△	○
Method E Distant consultancy	Yes	△	○

The results revealed a marginally significant effect of access to object of reference on memory retention ($F(1,22)= 3.063$, $p = .094$, $\eta_p^2 = .122$), a non-significant effect on case comprehension ($F(1,22) = 1.451$, $p = .241$) and a non-significant effect on subjective evaluation ($F(1,22) = 0.360$, $p =.555$) (Table 7-8). There were no interaction effects between visual access to object of reference and case complexity on any of the three measures (memory retention: $F (1,22) =.004$, $p = .952$; case comprehension: $F (1,22) = .000$, $p =1.0$; subjective evaluation; $F (1,22) =.022$, $p =.882$).

Table 7-8 Mean value of memory retention, case comprehension and subjective evaluation at each level of visual access to object of reference and case complexity

Access to object of reference	Case Complexity	Memory Retention (Out of 30)	Case comprehension (Out of 10)	Subjective evaluation
Low	Simple	20.3	6.9	5.9
	Complex	20.6	6.8	5.6
High	Simple	22.9	7.5	5.8
	Complex	23.1	7.3	5.4

7.2.4 Effect of visibility on handover quality

Due to the experimental design, it was not possible to analyze the effect of visibility following a similar mixed design. Thus, a two-way between-subject ANOVA was conducted, assuming the data points were independent observations.

Table 7-9 Experimental design for testing visibility

	Complex	Simple
Method B Cell phone handover	△	○
Method F Face-to-face handover	○	△

The results revealed non-significant effects of visibility on memory retention ($F(1,22)= 0.074$, $p = .789$; observed power = 0.058), on case comprehension ($F(1,22) = 0.004$, $p = .951$; observed power = 0.050) and on subjective evaluation ($F(1,22) = 1.067$, $p =.313$; observed power = 0.167), and non-significant effects of case complexity on memory retention ($F(3,22)= 0.028$, $p = .868$), on case comprehension ($F(3,22)=0.071$, $p =0.793$), and on subjective evaluation ($F(3,22) = 0.823$, $p = .374$).

Table 7-10 Mean value of memory retention, case comprehension and subjective evaluation at each level of visibility and case complexity

Visibility	Case Complexity	Memory Retention (Out of 30)	Case comprehension (Out of 10)	Subjective evaluation
Low	Simple	20.3	6.9	5.9
	Complex	20.6	6.8	5.6
High	Simple	21.2	6.9	5.7
	Complex	20.3	6.7	5.3

Because of the lack of significance of case complexity, the data were pooled over across the two levels of complexity and a paired sample t test was conducted. The results again revealed non-significant effect of visibility on memory retention (t(1,24) = -.275, p = .785, observed power = .058), on case comprehension (t(1,23) = .063, p = .950, observed power = .050) and on subject evaluation (t(1,23) = 1.056, p = .302, observed power = .173).

Table 7-11 Mean value of memory retention, case comprehension and subjective evaluation at each level of visibility

Visibility	Memory Retention (Out of 30)	Case comprehension (Out of 10)	Subjective evaluation
Low	20.4	6.8	5.7
High	20.7	6.8	5.5

The results from the ANOVA and t test failed to reject the null hypothesis, with the lowest significance level being 0.785. However, failing to reject the null does not necessarily mean rejection of the alternative. Therefore, a Bayesian t test was conducted and the Bayes factors were calculated (Rounder, Speckman, Sun, & Morey, 2009). The Bayes factor is an odd ratio which allows researchers to express preference for either the null hypothesis or the alternative. The Bayes factor for the respective values of t and N were calculated (http://pcl.missouri.edu/bf-one-sample) and the results showed that the odds ratios for preferring the null hypothesis versus the alternative is 6.15 for memory retention, 6.37 for case comprehension and 3.76 for subjective evaluation.

Bayes factors	Memory Retention	Case comprehension	Subjective evaluation
$\frac{Null\ hypothesis}{Alternative}$	6.15	6.37	3.76

7.2.5 Subjective evaluation of the four affordances

16 subjects evaluated subjectively the importance of the four affordances after completing the experiment. The results shown in Table 7-12 showed a significant

difference on the perceived importance between the four affordances (F(3,45) = 5.910, p<.01). A post-hoc pairwise comparison showed that the importance of visibility was significantly less than that of reviewability (mean difference = -1.5, p <.05), of interactivity (mean difference = -1.625, p <.01) and of access to object of reference (mean difference = -1.5, p <.05). There were no significant differences between the other three affordances.

Table 7-12 Subjective evaluation on the importance of the four affordances

Affordance	**Mean (SD)**
Reviewability	6.4 (0.8)
Interactivity	6.5 (0.8)
Access to object of reference	6.4 (1.0)
Visibility	4.9 (1.8)

7.3 Discussion

Study 4 aimed to test the effects of the four affordances on handover memory retention and comprehension, as well as their interaction effects with case complexity. Four main effects and three interaction effects are hypothesized. Below this section discusses how the results support these hypotheses.

H1a: Communication medium supporting higher reviewability leads to better memory retention and case comprehension.

The results provided support for the effect of reviewability on memory retention (Table 7-4). Past studies showed that when the communication load exceeded the capacity of working memory, communication accuracy (measured by readback error) using text-based display or redundant display was significantly higher than using auditory display (Helleberg & Wickens, 2003; Wickens & Gosney, 2003) (Table 2-13). Our results were consistent with the past findings, clearly demonstrating a redundancy gain.

However, the positive effect of reviewability was not found in case comprehension. The lack of significance could have resulted from two causes. First, redundant display provides twice as much “data” as a single modality display (Wickens et al., 2003). Perception of such data requires more cognitive resources and hence limits the working memory capacity for data interpretation and synthesis, which are essential components for comprehension. Second, it was highly likely that the content of the two modalities presented was not exactly the same. It was observed throughout Study 4 that the messages presented through the text-based channel were more formal and concise than those through the auditory channel. Therefore, the presentation or even the sequence of presentation of the auditory information may not synchronize with the eye fixation of each text-based message. Such a “fusing failure” (Wickens et al., 2011) could be quite disruptive, increasing the difficulty to synthesize and comprehend information, in a way that could offset the advantage of the redundancy of channels.

H1b: When case complexity is high, reviewability will more positively enhance memory retention and case comprehension.

The results of Study 4 did not provide statistical evidence for the interaction effect, despite a large difference of memory retention in complex cases (Difference = 3.8) versus simple cases (Difference = 1.9) (Table 7-4). The lack of significance could have resulted from the small sample size. It is likely that with a larger sample size an interaction effect would have been found.

H2a: Communication medium supporting higher interactivity leads to better memory retention and case comprehension.

The results supported the effect of interactivity on memory retention and case comprehension (Table 7-6). There is a great body of research documenting the positive advantage of interactivity - the ability of listeners to provide feedbacks and ask questions at the very moment when they experience communication breakdowns has been shown to be an important mechanism for achieving shared understanding (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Oviatt & Cohen, 1991) (Table 2-13). Our findings on case comprehension were consistent with these past studies, although no prior work has documented this important benefit in the context of medical handovers.

Moreover, our finding revealed the effect of interactivity on memory retention as well. The positive effect may have resulted from the handover receivers' ability of probing for more information using interactive channels. Past studies and our exploratory study showed that handover receivers prompt for additional information if not delivered voluntarily by the senders (Ye et al., 2007). The advantage of interactivity at ensuring adequate transfer of patient information was particularly important when the handover communication was unstructured or the handover participants were less experienced, for the handover senders in these cases may have higher possibility of omitting critical patient information.

H2b: When case complexity is high, interactivity will more positively enhance memory retention and case comprehension.

Similar to H1b, despite the larger performance differences in complex cases over simple cases (Memory retention: complex cases difference = 5.9, simple cases difference

= 2.0; case comprehension: complex cases difference = 1.6, simple cases difference = 1.2), the results of Study 4 did not support H2b (Table 7-6). The lack of significance could have been due to the small sample size.

H3a: Communication medium supporting visual access to object of reference leads to memory retention and case comprehension.

The results provided partial support for the effect of access to object of reference to memory attention (Table 7-8). Access to objects of reference for tasks involving reference of physical objects enables communication participants to view the same set of objects they are discussing. It helps the senders and the receivers understand the current state of their task and enable them to communicate and ground their conversations more effectively (Clark & Krych, 2004; Kraut et al., 1996; Whittaker et al., 1993) (Table 2-13). In clinical handovers, being able to view an ECG may help the receiver to perceive and understand the patient situation more effectively and efficiently than a message of “New T wave inversion on leads V2-V4”. In addition, seeing a patient painfully pressing his chest while receiving the message “sudden onset of chest pain” provides redundancy information which may have resulted in better information encoding.

Although the result of Study 4 did not provide statistical evidence for the effect of visual access to object of reference on case comprehension, it has to be noted that there was a performance difference between the two levels of manipulation (High access to objective of reference: 7.4; low access to objective of reference: 6.85). It is likely that with a larger sample size, a significant effect would have been found. Another possible explanation for the lack of significance is related to the limited working memory capacity, similar to the case of reviewability. It is possible that the redundant visual data may act as an "attention sink" (Wickens et al., 2003) competing for the working memory resources necessary for data interpretation and integration.

H3b: When case complexity is high, access to object of reference will more positively enhance memory retention and case comprehension.

Study 4 did not support the interaction effect between access to object of reference and case complexity. In addition, the performance differences between complex cases and simple cases were remarkably close (Memory retention: complex cases difference = 2.5, simple cases difference = 2.6; case comprehension: complex cases difference = 0.5, simple cases difference = 0.6) (Table 7-8). One possible explanation for this is due to the manipulation of complexity for access to object of reference. It was found posteriorly that case complexity (in terms of the amount of patient information and treatment difficulty) did not necessarily correspond to visual information complexity. A very complex case may be associated with clear chest X-ray and ECG, which indicates low visual information complexity. Therefore, in order to examine the interaction effect, more empirical studies with careful manipulation of visual information complexity are needed.

H4: Communication medium supporting visibility does not lead to better memory retention and case comprehension.

The results of Study 4 provided strong evidence for the null effect of visibility on memory retention and case comprehension (Table 7-10). The statistical significance of the ANOVA and t test were remarkably larger than 0.05, with the lowest p value being 0.785 (Table 7-11). In addition, the calculation of Bayes factors showed that there was a good amount of evidence in favor of the null hypothesis (A number larger than 10 is considered substantial (Rounder et al., 2009)). The null effect of visibility is further supported by the perceived low importance of visibility, compared to the other three affordances.

The result contradicts the opinions long-held by expert clinicians, that face-to-face handover is preferred to mediated communication (Solet et al., 2005). Being able to see each other did not improve the handover receivers' memory retention, case comprehension and subjective evaluation. The findings are of a highly practical implication for hospitals where the handover sender/receiver ratio is high. For example, the handover sender/receiver ratio in Singapore hospitals can be up to 20:1 (Yang et al., 2011), a few times higher than the figures reported in the US and Australia (Alem et al., 2008; Borowitz et al., 2008). A high sender/receiver ratio and the resulting large number of patients for each handover receiver impose more constraints on the coordination of handover time and location – face to face handover for every patient is virtually impossible.

It has to be noted that, Study 4 focused on the cognitive aspects of handover quality. Some studies posited that besides the primary objective of transferring critical patient information and care plan, handovers may serve the purpose of colleague bonding and medical education (Cohen & Hilligoss, 2010). In those cases, the exchange of emotions and altitudes is likely to be of high importance and visibility may influence the results.

Subjective measures and objective measures

The results of Study 4 showed that participants' subjective evaluations reflected relatively well the objective handover quality (Table 7-4, Table 7-6). The results is inconsistent with some past studies on situation awareness showing that subjective measures were not significantly correlated with objective ones (Endsley, Selcon, Hardiman, & Croft, 1998). The current findings may have resulted from multiple reasons. First, it is possible that, medical doctors, who receive frequent feedbacks through evidence based medicine, have better metacognitive abilities to judge the handover performance. Second, it was emphasized to the participants that the purpose of the study was not to test their clinical knowledge, but to compare the effectiveness of different handover methods. Thus, it is possible that the subject evaluation was perceived by the receivers as an evaluation of the senders' handover ability, instead of the receivers themselves, resulting in a more objective rating. Third, it has to be noted that the subject evaluation was collected as the last question in the questionnaire. The performance for objective memory retention and case comprehension questions may have helped the handover receivers to benchmark their performance and provide more accurate evaluation.

To sum up, Study 4 showed that the affordances of reviewability, interactivity and visual access to object of reference had positive effects on handover receivers' memory retention. The affordance of interactivity improved the receivers' case comprehension. In addition, Study 4 provided evidence for the null effect of the visibility - being able to see each other affected neither handover receivers' memory retention nor case comprehension. Similar to Study 3, Study 4 was limited in its fidelity of simulating the interaction with patients. Also, the sample size used in Study 4 was small, which might have contributed to some insignificant results in spite of evident trends. The takeaway points of Study 4 are summarized as follows.

- Communication medium supporting reviewability leads to better memory retention of patient information
- Communication medium supporting interactivity leads to better memory retention of patient information and care comprehension
- Communication medium supporting visual access to object of reference leads to better memory retention
- Communication medium supporting visibility does not lead to better memory retention nor case comprehension

8. Concluding remarks

8.1 General discussion

The continuity of patient care and limited resident working hours routinize the handover of patients between doctors. Ideally, handover receivers should have the same understanding of a patient as handover senders. In reality, however, clinical handovers were found to be vulnerable episodes contributing to near misses, adverse events and waste of healthcare resources (Horwitz et al., 2008).

Clinical handovers can be conceptualized as a three-phase process, namely pre-handover, handover communication, and post-handover (Kerr, 2002; Raduma-Tomàs et al., 2010). Since accurate and adequate handover preparation and communication are the prerequisites to ensure proper post-handover performance, this research focused on the first two phases, aiming to propose a suitable framework for modelling the factors influencing the performance of pre-handover and handover communication. The second objective was to study experimentally the effect of certain factors.

To achieve the research objectives, a series of literature review and four empirical studies were conducted. The literature review provided theoretical foundations for the research framework and research models, as well as revealed two main research gaps: the use of memory-intensive information access strategies at pre-handover and the lack of research with regard to handover media. Following the research framework, one exploratory study in a university hospital was conducted. The exploratory study employed semi-structured interview, recording of handover conversation and shadowing of handover receiver. The results of the exploratory study were consistent with the general findings from existing medical literature. First, medical residents depended largely on their memory when accessing patient information at the pre-handover phase. Second, the affordances of interactivity and reviewability, and case complexity influenced the information transfer during handover communication. Following the exploratory study, three experimental studies were carried out. Study 2 and 3 examined the relationship between information access cost, confidence judgment, overconfidence and information access attempts and task performance. Study 2 employed a general laboratory-based experimental paradigm with university students as subjects. Study 3 applied the same research model to examine medical residents' pre-handover performance, with residents from two tertiary hospitals as subjects. Study 2 and study 3 supported that both information access cost and confidence judgment influenced information access strategy and overconfidence affected task performance negatively. Study 4 was conducted to investigate the effects of four affordances on handover communication effectiveness.24 medical residents from the two universities participated in the experiment to hand over and receive 6 patients using 6 different handover methods. The results of study 4 supported the positive effects of interactivity, reviewability, and visual access to object of reference.

8.1.1 Information access strategy in general and in clinical pre-handovers

The exploratory study interviewed doctors on their pre-handover practices and found that the participants were remarkably confident in their memory of patient information and tended to use a memory-intensive strategy, in accordance with past studies (Alem et al., 2008; Vidyarthi et al., 2006). Following the exploratory study, Studies 2 and 3 were conducted to examine the relationship between information access costs (IAC), perceived accuracy (confidence judgment), information access behavior and performance (the model proposed in Figure 3-2). Utilizing two different methods to "manipulate" accuracy, confidence judgment and overconfidence, results from Study 2 supported the conclusion that besides the well-documented determinant IAC (Gray & Fu, 2001, 2004; Gray et al., 2006), the perceived accuracy of memory, or knowledge in the head (KIH), is another important determinant influencing the information access strategy. As confidence judgment increased, there were fewer information access attempts to knowledge in the world (KIW) such as printed documents (17% reduction in information access attempts). Furthermore, the accuracy-confidence inversion pattern (Tulving, 1981) in Study 2 offered strong support for the proposition that the perceived accuracy, rather than the actual accuracy (which showed a reversed pattern in Study 2), determined the information access strategy.

Consistent with past studies (Gray & Fu, 2001, 2004; Gray et al., 2006), IAC was found to affect performance in Study 3 – higher IAC resulted in poorer performance (16% reduction in accuracy). Moreover, although confidence judgment determines information access strategy, this dissertation argues that, overconfidence, rather than confidence judgment, affects performance (accuracy of answers in test 2). This is because under the condition of perfect calibration, where the confidence judgment reflects perfectly the actual accuracy, any changes in information access attempts due to varying confidence judgment can be considered as a result of adaptive decision making. Under the conditions of overconfidence, however, confidence judgment is overestimated, biasing the decision toward a KIH (memory) strategy more than it should and thereby impairing performance. The findings of Study 2 showed a significant main effect of overconfidence on performance. In addition, Study 3 showed the interaction effect between IAC and overconfidence on performance - as IAC increased overconfidence harmed the performance more negatively (10% reduction in accuracy). The result is consistent with the effort-performance space representation, stating that when resources become scarce, the highest utility is achieved through allocating more effort to activities which yield the highest expected gain (Wickens et al., 2004). Interestingly, the findings of Study 2 and 3 consistently showed that when IAC was low, participants made far more information access attempts to KIW that they should had done, even though their confidence judgment was high. The extra access attempts moderated the effect of overconfidence on performance.

With regard to clinical pre-handover, specifically, Study 3 clearly demonstrated that the increasing difficulty to access patient information harmed significantly the pre-handover preparation. Moreover, Study 3 showed that medical residents were overconfident in their memories, in both high and low accountability conditions. This overconfidence negatively influenced their pre-handover preparation accuracy when patient information was not easily accessible, a common circumstance found in most hospitals.

Findings from this research provide practical implications for improving medical resident's pre-handover accuracy. First, hospitals should try to reduce information access cost through every possible means, such as setting up central repositories to store patient information, and upgrading IT hardware to increase information retrieval speed. In particular, today many hospitals are trying to implement electronic handover tools, with the objective to increase patient safety (Flanagan et al., 2009). Preferably, such tools could be integrated with the Electronic Medical Records (EMR) so that a large amount of patient information could be extracted directly from the EMR – the information access cost from KIW becomes negligible in this case. Second, hospitals should try to improve medical residents' accuracy-confidence calibration. Viable solutions include education on the fallibility of human memory, handover training with constant feedback of the accuracy-confidence calibration, and increasing the accountability of pre-handover performance.

8.1.2 Communication media in clinical handovers

The findings of the exploratory study on handover communications provided qualitative and quantitative evidence for the importance of interactivity and reviewability. The ability to clarify doubts and ask questions, obtained by verbal communication was identified as a major benefit of the current handover practice in the university hospital. In addition, the analysis of information transfer between outgoing physicians (handover sender) and on-call physicians (handover receiver) showed that although outgoing physicians initiated the majority of the discussion, on-call physicians did prompt for more patient information. The outgoing physicians and the on-call physicians worked together to achieve a shared understanding of a patient.

In addition, the practice of mobile phone handover supported the affordance of semi-reviewability, where the handover receivers were allowed to record information at their own discretion. Even under the affordance of semi-reviewability, there was large information loss in the process. On average, 9.8 categories of information were discussed in a handoff but only 5.5 were recorded on paper by the on-call physicians (**Table 4-4**), resulting in a nearly 40% loss of handed over information.

Study 4 was carried out to study the effect of the identified four affordances and their interaction effects with case complexity (Figure 8-2). The results showed that reviewability (14% increase) and visual access to object of reference (12% increase) strongly supported the handover receivers' memory retention of important patient information, and interactivity supported both memory retention (24% increase) and case comprehension (26% increase). Furthermore, the results of Study 4 provided contradicting evidence to the long held recommendation that face to face handover is preferred to mediated communication (Solet et al., 2005). Being able to see other was found to affect neither memory retention nor case comprehension. Indeed there is a strong necessity for the medical community to unravel the affordances of visual access to object of reference and being able to see each other when making handover recommendations. It is possible that this perceived value of face-to-face situations is in fact reflective of the fact that most such situations also support a common visual access to the objects. In addition, interestingly, Study 4 showed that doctors' subjective evaluation of handover performance was relatively consistent with the objective results, in contrast to past results in Aviation (Endsley et al., 1998).

These findings offer practical implications for hospitals when implementing standard handover communication protocol. First of all, synchronous communication is preferred to asynchronous communication. Hospitals should provide the handover receiver a chance to ask questions and clarify doubts. Second, hospital could mandate the preparation of handover notes by the senders, with which full reviewability is afforded. It had to be noted, that the information access cost should be taken into consideration here because it influences the note preparation quality. Third, hospitals should create a "shared workplace", especially for complex cases, where the handover senders and receivers can view, or even manipulate the same set of patient information.

8.2 Limitations

This research has several limitations and challenges, which should be taken into consideration when interpreting and generalizing the findings. First of all, the mock up patients in Study 3 and 4 were simulated through patient photos, patient case files and diagnostic test results. The simulation was limited in its fidelity, since the interaction with a patient in real settings usually takes place through more modalities, such as through auditory and tactile channels, and usually lasts for a much longer time. The differences in interaction experiences may affect the pre-handover and handover communication performance. However, it has to be noted that pure text-based simulation has been used in past studies, generating satisfactory results (Bhabra et al., 2007; Pothier et al., 2005). Still, the simulation of patient cases for studying clinical handovers has not received much attention from the research community and definitely more empirical studies are needed to determine the appropriate fidelity.

Second, the sample size used in Study 4 was small, contributing to some insignificant results in spite of evident trends. As the research aimed to investigate the clinical handover performance, it was crucial to engage participants who had experiences with the task. Unfortunately, medical residents are not easy to recruit, due to their busy schedules and other high priority tasks. In fact, 6 participants who joined Study 3 dropped out of Study 4 because it was difficult to schedule a timeslot at which both participants in a pair were available.

Third, the participants of Study 3 and 4 were medical residents, with limited clinical experience. The findings from the studies are limited in external validity. As described in the literature review, experience and training affect one's accuracy-confidence calibration. More experienced clinicians are likely to possess a smaller overconfidence bias. More empirical research is needed to generalize the research results to shift handovers between experienced clinicians.

Fourth, Study 2 and Study 3 might have overestimated the tendency to access information from the world, as the two studies were framed in a single task performance paradigm. However, studies have shown that handover is one of many concurrent tasks that must be managed by incoming and outgoing clinicians (Abraham, Kannampallil, & Patel, 2014; Carroll et al., 2012). These concurrent tasks would highly likely affect the performance of the information retrieval behavior and performance. It is likely that the cost of access knowledge in the world would increase and the outgoing doctor would be more inclined to depend on their memory.

Fifth, in study 4, the different communication media conditions were manipulated within-subject with a randomized presentation sequence. Although it was emphasized that the participants should follow their daily practices of shift handovers, a carry-over effect was still possible.

8.3 Future direction

This research sheds light on several future directions. With regard to pre-handover performance, it would be interesting to examine how the feedback loop affects subjects' accuracy-confidence calibration, which in turn influences the information access behaviors and performance. Fennema and Kleinmuntz (1995) found that the decision makers had only limited ability to anticipate effort accuracy attained for given tasks, even after receiving feedbacks. Would similar results be found with doctors at pre-handover? Another exciting area worth investigating for the pre-handover phase is the reason why participants choose to refer to KIW more than necessary when IAC is low. Is this strategy unique to Asians (mainly Chinese ethnics) or is it a universal trait?

With regard to communication media, the first extension of study could be to examine the "fit" between handover participants and handover media. According to Cohen and Hilligoss (2009), handovers could be categorized into three broad categories: a continuing patient transfer, a new patient transfer, and a cross-boundary transfer. In each kind of handovers, the participants have varying extent of shared knowledge, expertise and practice, and the degree to which handover receivers know a patient to be handed over (Coiera, 2000; Nemeth et al., 2008). How does such "fit" between handover participants and media affect the handover outcomes? In addition, this research did not examine the interaction effects between different affordances. For example, studies have suggested that allowing communication participants to see each other and the shared workplace at the same produces better communication outcomes than allowing them either to see each other or to view the same workplace (Watts & Monk, 1996). More empirical studies are needed to examine the interaction effects in handover communication.

The second extension of research on handover communication is to incorporate the socioemotional/relational dimension. The focus of this thesis lies in the cognitive dimension of handover communication. However, as pointed out in several articles, that the purpose of handover is not limited to accurate transfer of information, but include other functions such as learning and colleague bonding (Abraham et al., 2014; Carroll et al., 2012; Cohen & Hilligoss, 2010; Wears, 2012), where the socioemotional/relational dimension plays a large role. Incorporation of this dimension would provide a more comprehensive measure of handover effectiveness. Also, according to the common ground theory, communication parties are shaping the communication processes and outcomes by minimizing their collaborative effort (Clark & Brennan, 1991). Hence, a closer look into the communication process would be a nice welcome.

Another potential research direction for both the pre-handover phase and the handover communication phase is to analyze the handover process within a broader clinical context. Study 1 of this thesis revealed that the incoming doctors were constantly performing other tasks while receiving handover calls. Other studies have also noted that handover is one of many concurrent tasks that must be managed by incoming and outgoing clinicians (Abraham et al., 2014; Carroll et al., 2012). These concurrent tasks would highly likely affect the performance of the primary handover task.

In addition, while it may be better to have further validation of the design implications, funding resources and the incorporation of a tool/system into the hospital processes are difficult to implement within the research time period. More empirical studies are needed to validate any handover tools developed based on the results of this research.

Reference

Abraham, J., Kannampallil, T., & Patel, V. L. (2014). A systematic review of the literature on the evaluation of handoff tools: implications for research and practice. *Journal of the American Medical Informatics Association: JAMIA, 21*(1), 154-162.

ACQSHC. (2009). *The OSSIE Guide to Clinical Handover Improvement*, Sydney.

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research, 27*, 123-157.

Alem, L., Joseph, M., Kethers, S., Steele, C., & Wilkinson, R. (2008). Information environments for supporting consistent registrar medical handover. *HIM Journal, 37*(1), 9-25.

Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness judgments: The effect of dyadic collaboration. *Applied Cognitive Psychology, 17*, 545-561.

Anderson, C. D., & Mangino, R. R. (2006). Nurse shift report: who says you can't talk in front of the patient? *Nursing Administration Quarterly, 30*(2), 112-122.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Arora, V., Johnson, J., Lovinger, D., Humphrey, H. J., & Meltzer, D. O. (2005). Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis. *Quality and Safety in Health Care, 14*(6), 401-407.

Austin, J. (1960). *How to do things with words*. Oxford: Clarendon Press.

Ballard, D. H., Hayhoe, M., M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723-767.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66-80.

Bavelas, J., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952.

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes, 45*, 111-139.

Bhabra, G., Mackeith, S., Monteiro, P., & Pothier, D. (2007). An experimental comparison of handover methods. *Annals of the Royal College of Surgeons of England, 89*(3), 298-300.

Blais, A., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. *Personality and Individual Differences, 38*, 1701-1713.

Borgida, E., & Howard-Pitney, B. (1983). Personal involvement and the robustness of perceptual salience effects. *Journal of Personality and Social Psychology, 34*, 978-989.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of experimental psychology: Applied, 5*(1), 76-88.

Borowitz, S. M., Waggoner-Fountain, L. A., Bass, E. J., & Sledd, R. M. (2008). Adequacy of information transferred at resident sign-out (in-hospital handover of care): a prospective survey. *Quality and Safety in Health Care, 17*(1), 6-10.

Brennan, S. E., & Lockridge, C. B. (2006). Computer-mediated communication: a cognitive science approach In K. Brown (Ed.), *ELL2*, *Encyclopedia of Language and Linguistics*, *2nd Edition* (pp. 775-780). Oxford, UK: Elsevier Ltd.

Carroll, J. S., Williams, M., & Gallivan, T. M. (2012). The ins and outs of change of shift handoffs between nurses: a communication challenge. *BMJ Quality & Safety, 21*, 586-593.

Casti, J. L. (1979). *Connectivity, Complexity and Catastrophe*. New York: Wiley.

Chapanis, A., Ochsman, R., Parrish, R., & Weeks, G. (1972). Studies in interactive communication: I The effects of four communication modes on the behavior of teams during cooperative problem solving. *Human Factors, 14*, 487-509.

Chapanis, A., Ochsman, R., Parrish, R., & Weeks, G. (1977). Studies in interactive communication. II: The effects of four communication modes on the linguistic performance of teams during cooperative

problem solving. *Human Factors, 19*(2), 101-126.

Chen, P. W. (2009, September 3, 2009). When Patient Handoffs Go Terribly Wrong, *The New York Times*. Retrieved from http://www.nytimes.com/2009/09/03/health/03chen.html?_r=0

Chu, P. C., & Spires, E. E. (2003). Perceptions of accuracy and effort of decision strategies. *Organizational Behavior and Human Decision Processes, 91*, 203-214.

Clark, H. H. (1996). *Using Language*. Cambridge: CUP.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC.: American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62-81.

Clark, H. H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-292.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *cognition, 22*, 1-39.

Cohen, M. D., & Hilligoss, P. B. (2009). Handoffs in Hospitals: A review of the literature on information exchange while transferring patient responsibility or control. http://deepblue.lib.umich.edu/handle/2027.42/61498

Cohen, M. D., & Hilligoss, P. B. (2010). The published literature on handoffs in hospitals: deficiencies identifies in an extensive review. *Quality and Safety in Health Care, 19*, 493-497.

Coiera, E. (2000). When conversation is better than computation. *Journal of the American Medical Informatics Association, 7*(3), 277-286.

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physics Review Letters, 63*(105).

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial information processing and organization design. In B. Staw & L. Cummings (Eds.), *Research in Organizational Behavior* (pp. 191-233). Greenwich, Connecticut: JAI Press.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message equivocality, media selection and manager performance: implications for information systems. *MIS Quarterly, 11*(3).

Dawis, R. V., Lofquist, L. H., & Weiss, D. J. (1968). A theory of work adjustment: A revision. *Minnesota Studies in Vocational Rehabilitation, 23*, 15.

Dennis, A. R., Kinney, S., & Hung, Y. (1999). Gender differences in the effects of media richness. *Small Group Research, 30*(4), 405-437.

Dennis, A. R., & Valacich, J. S. (1999). *Rethinking media richness: Towards a theory of media synchronicity.* Paper presented at the the 32nd Hawaii International Conference on System Sciences, Maui, Hawaii.

Dennis, A. R., Wixom, B. H., & Vandenberg, R. J. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis. *MIS Quarterly, 25*, 167-193.

DeVoge, J. M., Bass, E. J., Atia, M., Bond, M., Waggoner-Fountain, L., & Borowitz, S. M. (2009). *The Development of a Web-based Resident Sign-out Training Program.* Paper presented at the IEEE International Conference on Systems, Man and Cybernetics.

Dobbins, I. G., Kroll, N. E. A., & Liu, Q. (1998). Confidence-accuracy inversions in scene recognition: A remember-know analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1306-1315.

Dunning, D., & Story, A. L. (1991). Depression, realism, and the overconfidence effect: Are the sadder wiser when predicting future actions and events. *Journal of Personality and Social Psychology, 61*(4), 521-532.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 85-104.

Endsley, M. R., & Jones, W. M. (2001). A model of inter-and intrateam situation awareness: implications for design, training, and measurement. In M. McNeese, E. Salas & M. R. Endsley (Eds.), *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments* (pp. 46-67). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). *A comparative analysis of SAGAT and SART for evaluations of situation awarenss.* Paper presented at the the Human Factors and Ergonomics Society 42nd Annual Meeting Chicago, IL.

Fennema, M. G., & Kleinmuntz, D. N. (1995). Anticipations of Effort and Accuracy in Multiattribute Choice. *Organizational Behavior and Human Decision Processes, 63*(1), 21-32.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564.

Flanagan, M. E., Patterson, E. S., Frankel, R. M., & Doebbeling, B. N. (2009). Evaluation of a Physician Informatics Tool to Improve Patient Handoffs *Journal of the American Medical Informatics Association, 16*(4), 509-515.

Goodhue, D. L. (1995). Understanding user evaluation of information systems *Management Science, 41*(12), 1827-1844.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quaterly, 19*(2), 213-237.

Granhag, P. A., Stromwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindersight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology, 14*, 397-420.

Gray, W. D., & Fu, W.-T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. *CHI Letters, 3*(1), 112 - 119

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*(3), 359-382.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 133*(3), 461-482.

Gudykunst, W., & Kim, Y. (1992). *Communicating with strangers: An approach to intercultural communication* (2nd ed.). New York: McGraw-Hill.

Harvey, C. M., Schuster, R. J., Durso, F., T., Matthews, A. L., & Surabattula, D. (2006). Human Factors of Transition of Care. In P. Carayon (Ed.), *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety*: Lawrence Erlbaum Associates.

Helleberg, J., & Wickens, C. D. (2003). Effects of data-link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology., 13*(3), 489-210.

Henrion, M., & Fischhoff, B. (2002). Assessing uncertainty in physical constants. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 666-677). NY: Cambridge University Press.

Hilligoss, P. B., & Zheng, K. (2013). Chart biopsy: an emerging medical practice enabled by electronic health records and its impacts on emergency department-inpatient admission handoffs. *Journal of American Medical Informatics Association, 20*, 260-267.

Hiltz, S. R., Johnson, K., & Turoffm, M. (1986). Experiments in group decision making communication process and outcome inface-to-face versus computerized conferences. *Human Communication Research, 13*(2), 225-252.

Hollins, T. S., & Perfect, T. J. (1997). The confidence-accuracy relation in eyewitness memory: The mixed question type effect. *Legal and Criminological Psychology, 2*, 205-218.

Horwitz, L. I., Moin, T., Krumholz, H. M., Wang, L., & Bradley, E. H. (2008). Consequences of inadequate sign-out for patient care. *Archives of Internal Medicine, 168*(16), 1755-1760.

Horwitz, L. I., Parwani, V., Shah, N. R., Schuur, J. D., Meredith, T., Jenq, G. Y., & Kulkarni, R. G. (2009). Implementation and evaluation of an asynchronous physician sign-out for emergency department admissions. *Ann Emerg Med, 54*, 368-378.

Jarvenpaa, S. L., Knoll, K., & Leidner, D. E. (1998). Is Anybody Out There? Antecedents of Trust in Global Virtual Teams. *Journal of Management Information Systems Research, 14*(4), 29-64.

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed.). Boston: Pearson Education, Inc.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*(2), 177-192.

Joint Commission. (2006). 2006 Critical Access Hospital and Hospital National Patient Safety Goals. Retrieved January 07, 2013, from http://www.jointcommission.org/PatientSafety/NationalPatientSafetyGoals/06_npsg_cah.htm

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D. (2011). *Thinking, Fast and Slow* (1st ed.): Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes, 39*, 98-114.

Kerr, M. P. (2002). A qualitative study of shift handover practice and function from a socio-technical perspective. *Journal of Advanced Nursing, 37*, 123-134.

Kitch, B. T., Cooper, J. B., Zapol, W. M., Marder, J. E., Karson, A., Hutter, M., & Campbell, E. G. (2008). Handoffs causing patient harm: a survey of medical and surgical house staff. *Joint Commission Journal on Quality and Patient Safety, 34*(10), 563-570.

Koehler, D., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 686-715). NY: Cambridge University Press.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). Cambridge, UK: Cambridge University Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6*, 107-118.

Kraut, R., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 138-162). Cambridge, MA: MIT Press.

Kraut, R., Gergle, D., & Fussell, S. R. (2002). *The use of visual information in shared visual spaces: Informming the development of virtual co-presence.* Paper presented at the CSCW'02, New Orleans, Louisiana, USA.

Kraut, R., Lewis, S., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology, 43*, 718-731.

Kraut, R., Miller, M. D., & Siegel, J. (1996). *Collaboration in performance of physical tasks: effects on outcomes and communication.* Paper presented at the the 1996 ACM conference on Computer supported cooperative work

Krug, K. S., & Weaver, C. A. I. (2005). Eyewitness memory and product identification:

Evidence for familiarity biases. *Journal of General Psychology, 132*(4), 429-445.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees:How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes, 47*, 567-587.

Lamond, D. (2000). The information content of the nurse change of shift report: a comparative study. *Journal of Advanced Nursing, 31*(4), 74-804.

Lawrence, R. H., Tomolo, A. M., Garlisi, A. P., & Aron, D. C. (2008). Conceptualizing handover strategies at change of shift in the emergency department: a grounded theory study. *BMC Health Services Reserch, 8*, 256-267.

Lazarsfeld, P. F., & Merton, R. K. (1964). Friendship as social process: a substantive and methodological analysis. In M. Berger (Ed.), *Freedom and Control in Modern Society*. New York: Octagon.

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: the critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care, 13*(Suppl 1), 85-90.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes, 20*, 159-183.

Lichtenstein, S., & Fischhoff, B. (1980). How well do probability experts assess probability? *Decision Research Report*.

Lichtenstein, S., & Fischhoff, B. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge, England: Cambridge University Press.

Lu, S., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors*.

Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology, 92*(1), 152-159.

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident, but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*, 114-121.

McDonald, S. (2005). Studying actions in context: a qualitative shadowing method for organizational research *Qualitative Research, 5*(4), 455-473.

MOH Holdings. (2014). *Medical service scheme*. Retrieved from http://www.physician.mohh.com.sg/medical_service_scheme.html

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving Memory after Interruption: Exploiting Soft Constraints and Manipulating Information Access Cost. *Journal of experimental psychology: Applied, 15*(4), 291-306.

Morgan, P. L., Waldron, S. M., King, S. L., & Patrick, J. (2007). *Harder to Access, Better Performance? The Effects of Information Access Cost on Strategy and Performance.* Paper presented at the HCII 2007.

Morrow, D., & Fisher, U. M. (unknown). *Communication in Socio-Technology Systems*. New York, NY: Oxford University Press.

Murphy, A. H., & Daan, H. (1984). Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zierikzee experiment. *Monthly Weather Review, 112*, 413-423.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-26). Cambridge, MA: MIT Press.

Nemeth, C. P., Kowalsky, J., Brandwijk, M., Kahana, M., Klock, P. A., & Cook, R. I. (2008). Between Shifts: Healthcare Communication in the PICU. In C. P. Nemeth (Ed.), *Improving Healthcare Team Communication: Building on Lessons from Aviation and Aerospace* (pp. 135-153): Ashgate.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newlands, A., Anderson, A. H., & Mullin, J. (2003). Adapting communicative strategies to computer-mediated communication: An analysis of task performance and dialogue structure. *Applied Cognitive Psychology, 17*, 325-348.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Norton, R. (1983). *Communicator style: Theory, application and measures*. Newbury Park, CA: Sage.

Obstfelder, A., & Moen, A. (2006). The electronic patient record in community health services--paradoxes and adjustments in clinical work. *Studies in Health Technology and Informatics, 122*, 626-631.

Olson, J. S., Olson, G. M., & Meadar, D. K. (1995). *What mix of video and audio is useful for small groups doing remote real-time design work?* Paper presented at the 1995 Conference on Human Factors in Computing Systems, Denver, Colorado, USA.

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology, 85*, 504-511.

Oskamp, s. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*(3), 261-265.

Oviatt, S., & Cohen, P. (1991). Discourse Structure and Performance Efficiency in Interactive and Non-Interactive Spoken Modalities. *Computer Speech and Language, 5*, 297-326.

Parasuraman, R. (1986). Vigilance, monitoring and search. In J. R. Boff, L. Kaufmann & J. P. Thomas (Eds.), *Handbook of Human Perception and Performance* (Vol. II, pp. 43.41-43.39). New York: Wiley.

Patterson, E. S., Roth, E. M., Woods, D. D., Chow, R., & Gomes, J. O. (2004). Handoff strategies in settings with high consequences for failure: lessons for health care operations. *International Journal for Quality in Health Care, 16*(2), 125-132.

Patterson, E. S., & Wears, R. L. (2010). Patient handoffs: Standardized and reliable measurement tools remain elusive. *The Joint Commission Journal on Quality and Patient Safety, 36*(2), 52-61.

Patterson, E. S., Weiss, C., Woods, Z., Mount-Campbell, A. F., & Rayo, M. J. (2013). *Online training for resilient communication during shift change handovers.* Paper presented at the the 57th Human Factors and Ergonomics Society Annual Meeting.

Payne, J. W., Bettman, J. R., & Johansson, M. (1993). The adaptive decision maker. New York: Cambridge University Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 85-92.

Philibert, I. (2009). Use of strategies form high-reliability organisations to the patient hand-off by resident physicians: practical implications. *BMJ Quality & Safety, 18*, 261-266.

Pothier, D., Monteiro, P., Mooktiar, M., & Shaw, A. (2005). Pilot study to show the loss of important data in nursing handover. *British Journal of Nursing, 14*(20), 1090-1093.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences, 23*(1), 125-133.

Raduma-Tomàs, M. A., Flin, R., Yule, S., & Close, S. (2010, 2010, September 27-October 1, 2010). *Doctors' handovers in an acute medical assessement unit: A hierarchical task analysis.* Paper presented at the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA.

Randell, R., Wilson, S., & Woodward, P. (2011). The importance of the verbal shift handover report: a multi-site case study. *International Journal of Medical Informatics, 80*, 803-812.

Reid, A. (1977). Comparing the telephone with face to face interaction. In I. Pool (Ed.), *The Social Impact of the Telephone* (pp. 386-414). Cambridge, MA: MIT Press.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40*, 193-218.

Rounder, J. N., Speckman, P. L., Sun, D., & Morey, R. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Rutter, D. R. (1987). *Communication by telephone.* . Oxford, UK: Pergamon.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: implications for measurement and training. *Human Factors, 37*, 123-136.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication* Urbana: University of Illinois Press.

Short, J. (1972). Effects of medium of communication on experimental negotiation. *Human Relations, 27*, 225-234.

Singh, H., Thomas, E. J., Petersen, L. A., & Studdert, D. M. (2007). Medical errors involving trainees: a study of closed malpractice claims from 5 insurers. *Archives of Internal Medicine, 167*(19), 2030-2036.

Sledd, R., Bass, E. J., Borowitz, S., & Waggoner-Fountain, L. (2006, October 16-20, 2006). *Characterization of sign-out in pediatric acute care wards to inform process improvement.*

Paper presented at the The 50th Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, CA.

Sniezek, J. A., Paese, P. W., & Switzer III, F. S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264-282.

Solet, D. J., Norvell, J. M., Rutan, G. H., & Frankel, R. M. (2005). Lost in Translation: Challenges and Opportunities in Physician-to-Physician Communication During Patient Handoffs. *Academic medicine, 80*(12), 1094-1099.

Stelter, E. M., & Wickens, C. D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks. *Human Factors, 48*(1), 166-181.

Stephenson, G., Ayling, K., & Rutter, D. (1976). The role of visual communication in social exchange. *British Journal of Social and Clinical Psychology, 15*, 113-120.

Strong, D. M., Dishaw, M. T., & Bandy, D. B. (2006). Extending task technology fit with computer self-efficacy. *ACM SIGMIS Database, 37*(2-3).

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2011). Prediction in situation awareness: Confidence bias and underlying cognitive abilities. *The International Journal of Aviation Psychology., 21*(2), 153-174.

Sutcliffe, K. M., Lewton, E., & Rosenthal, M. M. (2004). Communication failures: an insidious contributor to medical mishaps. *Academic medicine, 79*(2), 186-194.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52*(4), 700-709.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20*, 479-496.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research, 2*(1), 63-84.

Vidyarthi, A. R., Arora, V. M., Schnipper, J. L., Wall, S. D., & Wachter, R. M. (2006). Managing discontinuity in academic medical centres: strategies for a safe and effective resident sign-out. *Journal of Hospital Medicine, 1*(4), 257-266.

Wallsten, T. S. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making, 1*, 145-164.

Walther, J. B., Slovacek, C. L., & Tidwell, L. C. (2001). Is a Picture Worth a Thousand Words? Photographic Images in Long-Term and Short-Term Computer-Mediated Communication. *Communication Research, 28*(1), 105-134.

Watts, L., & Monk, A. (1996). *Remote assistance: a view of the work and a view of the face?* Paper presented at the CHI '96 Conference Companion on Human Factors in Computing Systems.

Wears, R. L. (2012). Poverty amid plenty. *BMJ Quality & Safety, 21*(7), 533-534.

Weisband, S. (1992). Group discussion and first advocacy effects in computer-mediated and face-to-face decision making groups. *Organizational Behavior and Human Decision Processes, 53*(352-380).

Welch, S. J., Cheung, D. S., Apker, J., & Patterson, E. S. (2013). Strategies for improving communication in the emergency department: Mediums and messages in a noisy environment. *The Joint Commission Journal on Quality and Patient Safety, 39*(6), 279-286.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review, 4*(3), 387-392.

Whittaker, S. (2002). Theories and methods in mediated communication. In A. Graesser, M. Gernsbacher & S. Goldman (Eds.), *The handbook of discourse processes* (pp. 243-286). Hillsdale, NJ: Erlbaum.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspace: how do they work and when are they useful? *International Journal of Man-Machine Studies, 39*, 813-842.
Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors, 37*(3), 473-494.
Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.
Wickens, C. D., & Gosney, J. L. (2003). *Redundancy, modality, and priority in dual task interference.* Paper presented at the Human Factors and Ergonomics Society 47th Annul Meeting, Denver, Colorado.
Wickens, C. D., & Hollands, J. G. (1999). *Engineering Psychology and Human Performance*: Prentice Hall.
Wickens, C. D., Hollands, J. G., Parasuraman, R., & Banbury, S. (2013). *Engineering Psychology and Human Performance* (4th ed.): Pearson.
Wickens, C. D., Lee, J. D., Liu, Y., & Becker, S. E. G. (2004). *An introduction to human factors*

engineering (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
Wickens, C. D., & MaCarley, J. S. (2007). *Applied Attention Theory*: CRC Press.
Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2011). *Auditory-visual redundancy in vehicle control interruptions: Two meta-analyses.* Paper presented at the Human Factors and Ergonomics Society 55th Annul Meeting, Las Vegas, Nevada, .
Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication. *Psychological Bulletin, 16*(963-976).
Wogalter, M. S., & Laughery, K. R. (2006). Warnings and Hazard

Communications. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 889-911). Hoboken, NJ: Wiley.
Yang, X., Koh, R. Y. I., Tay, B. T. C., Siah, K. T. H., Donchin, Y., & Park, T. (2011, September 19-23). *Clinical shift handoffs in Singapore: A three-phase prospective.* Paper presented at the Human Factors and Ergonomics Society 55th Annual Meeting, Las Vegas.
Yang, X., Park, T., Siah, K. T. H., Ang, S. B. L., & Donchin, Y. (2014). One size fits all? Challenges faced by physicians during shift handovers in a hospital with high sender/recipient ratio. *Singapore Medical Journal*.
Ye, K., McD Taylor, D., Knott, J. C., Dent, A., & MacBean, C. E. (2007). Handover in the emergency department: deficiencies and adverse effects. *Emergency Medicine Australasia, 19*(5), 433-441.
Zigurs, I., & Buckland, B. K. (1998). A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly, 22*(3), 313-334.

Appendix

Appendix A: Informed consent for study 1-4

NANYANG TECHNOLOGICAL UNIVERSITY

INFORMED CONSENT

National University Hospital

Clinical Handover Project for Decreasing Handover Errors

Interview on Clinical Handovers

This interview aims to understand the current handover practices between doctors. Results from this study will be used to guide future developments on improving handovers in hospitals. We are grateful for your contributions and inputs should you choose to participate in this study.

The interview will be conducted by two human factors specialists from Nanyang Technological University. Participants will be asked to describe their current practices in clinical handover. The interview takes approximately 30 minutes to complete. The conversation will be audio recorded for further analysis.

All documented information will be reviewed by the Medical Affairs Department to maintain patient confidentiality, and any sensitive data such as patient identity will be removed prior to research access. Any inputs made by participants will be kept anonymous.

There are no known physical or psychological risks towards participants. All information will be kept strictly confidential and used solely for the purpose of this study. Information irrelevant to this study will not be documented. There are no right or wrong answers to the questions asked and your performance will not be evaluated. Participants are free to stop and withdraw anytime during the course of this study.

If you have any queries or concerns, please do not hesitate to ask any of the researchers on site, or contact the principle investigator, Prof. Park Taezoon, at 6790-5087 or tzpark@ntu.edu.sg.

SIGNATURE : ______________________________

NAME : ______________________________

DATE : ______________________________

INFORMED SHEET

Clinical Handover Project for Decreasing Handover Errors

Understanding Clinical Handovers

This research aims to understand the requirements and cognitive processes involved during clinical handovers between doctors, as well as the strengths and weaknesses of handover practices currently being used. Results from this study will be used to guide future developments on improving handovers in hospitals. We are grateful for your contributions and inputs should you choose to participate in this study.

For this study, participants can contribute either when they are on duty as the MO3, or when they are in communication with the MO3 to hand over shift-related information. The participating MO3 will be shadowed by a researcher, who will be observing the MO3's tasks during the transition period from day to evening shifts. Communication between participating doctors and the MO3 during handovers, including those made over telephone calls, will be observed and/or recorded. Each period of shadowing and recording would last approximately 90 minutes, involving only the earlier portion of MO3's shift when handovers typically take place.

Participating MO3s would be equipped with a handphone during the scheduled shadowing period, which has the ability to record two-way conversations whenever the user chooses to do so. Throughout the course of the observation, the MO3 would go about with the daily routines, and may choose to provide inputs for the researcher whenever convenient. After each handover session, the MO3 would be polled briefly for comments about the handover.

Participating day-shift MOs may have their voices recorded when communicating with the MO3 being shadowed. Only shift-related conversations and information will be documented, and all other matters will be ignored or omitted.

All documented information will be reviewed by the Medical Affairs Department to maintain patient confidentiality, and any sensitive data such as patient identity will be removed prior to research access. The communications between out-going MOs and the MO3, the post-handover comments, and any inputs made by participants will be kept anonymous.

There are no known physical or psychological risks towards participants. All information will be kept strictly confidential and used solely for the purpose of this study. Information irrelevant to this study will not be documented. There are no right or wrong answers to the questions asked and your performance will not be evaluated. Participants are free to stop and withdraw anytime during the course of this study.

If you have any queries or concerns, please do not hesitate to ask any of the researchers on site, or contact the principle investigator, Prof. Park Taezoon, at 6790-5087 or tzpark@ntu.edu.sg.

SIGNATURE : ______________________________

NAME : ______________________________

DATE : ______________________________

NANYANG TECHNOLOGICAL UNIVERSITY

INFORMATION SHEET

Understanding the information retrieval behaviour in 'Read and Answer'

This study aims to investigate the interactive behaviors between you and the 'Study and Recognize' program and how it influences your performance. Results from this study will provide important insights into the designs of information support tools that will have wide applications in HCI fields. We would be extremely grateful for your contributions should you choose to participate in this study.

During the study, you will undergo multiple trials, using the 'Study and Recognize' program. In each trial, you will be asked to review some photos, perform a cognitive task and later on recognize the photos you have viewed before from some distractors.

You will not be subjected to any physical risks throughout the study. You are expected to be involved in the study for approximately 90 minutes. There will be adequate rest breaks and refreshments provided during the duration of the study. On completion of all the, you will be given S$15 in cash as reimbursement for the time participating in the study. All your identification information and data gathered from this study will be kept strictly confidential and used solely for the purpose of this study.

While your participation would be greatly appreciated, we would like to emphasize that participation in this study is completely voluntary and that you have the right to refuse involvement or to stop and withdraw from the study at any time, without needing to give reasons, by informing the researchers on site or the principle investigator (contact details below). All collected data from you will be discarded should you choose to withdraw from the study. We would also like to emphasize that any performance assessments made during the study only has relevance within the study and will not have any other implications beyond it.

If you have any queries or concerns, please do not hesitate to ask any of the researchers on site, or contact the principle investigator, Asst Prof. Taezoon Park at tzpark@ntu.edu.sg. For an independent opinion regarding the research and the rights of research participants, you may contact the Nanyang Technological University Institutional Review Board at IRB@ntu.edu.sg.

Thank you.

Please ***tick*** *accordingly:-*

- ☐ I have read and understood the information provided in the Participant Information Sheet.

- ☐ I acknowledge that I can withdraw from this study at any point in time by informing the researchers on site or the Principle Investigator and all my data will be discarded.

- ☐ My signature below is my acknowledgement that I have agreed to participate in this study according to the above-mentioned terms.

SIGNATURE : ______________________________

NAME : ______________________________

DATE : ______________________________

INFORMATION SHEET

Developing a handover tool for reducing physicians' workload during handover preparation

The study aims to investigate the factors that influence physicians' information collection behaviours at pre-handover phase. Our goal is to understand how you gather patient information for handover and propose design guidelines for handover tool development. The focus is not on your clinical abilities at all. Instead, we are interested in developing good technology interventions to assist your work. We are grateful for your contributions and inputs should you choose to participate in this study.

The study will be conducted by one human factors specialist from Nanyang Technological University (NTU). In the study, you will role play as a primary team physician - study some medical cases commonly found in internal medicine and later on prepare for the handover of these cases.

There are no known physical or psychological risks towards participants. All information will be kept strictly anonymous and confidential, and used solely for the purpose of this study. Information irrelevant to this study will not be documented. You are free to stop and withdraw anytime during the course of this study.

If you have any queries or concerns, please do not hesitate to ask any of the researchers on site, or contact the investigators, Dr. Kewin Siah Tien Ho at 98411151 (kewinsiah@nuhs.com.sg) or Asst Prof. Park Taezoon, at 6790-5087 (tzpark@ntu.edu.sg).

*Please **tick** accordingly:-*

- ☐ I have read and understood the information provided in the Participant Information Sheet.

- ☐ I acknowledge that I can withdraw from this study at any point in time by informing the researchers on site or the Principle Investigator and all my data will be discarded.

- ☐ My signature below is my acknowledgement that I have agreed to participate in this study according to the above-mentioned terms.

NAME : ____________________

SIGNATURE : ____________________

DATE : ____________________

DEMOGRAPHICS

Age: ____________ **Years after MBBS:** ____________________

Department: ____

INFORMATION SHEET

Developing a handover tool for fulfilling physicians' communication needs during handover

Hospitals adopts a variety of handover methods including face to face handover, handover at bedside, phone handover, formal sign-out notes, audio recording and etc. What is the most appropriate method for physicians in Singapore? This study aims to answer this question by studying the relationship between handover methods and handover outcomes. Results from the study will be used to guide the development of better handover tools. The focus is not on your clinical abilities at all. Instead, we are interested in developing good technology interventions to assist your work. We are grateful for your contributions and inputs should you choose to participate in this study.

The study will be conducted by one human factors specialist from Nanyang Technological University (NTU). In the study, you will role play as primary team physicians and on-call physicians. As the primary team physician, you will study some medical cases commonly found in internal medicine and later on use different methods to handover these cases to the on-call physicians. As the on-call physician, you will receive patients using different methods as well.

There are no known physical or psychological risks towards participants. All information will be kept strictly anonymous and confidential, and used solely for the purpose of this study. Information irrelevant to this study will not be documented. You are free to stop and withdraw anytime during the course of this study.

If you have any queries or concerns, please do not hesitate to ask any of the researchers on site, or contact the investigators, Dr. Kewin Siah Tien Ho at 98411151 (kewinsiah@nuhs.com.sg) or Asst Prof. Park Taezoon, at 6790-5087 (tzpark@ntu.edu.sg).

*Please **tick** accordingly:-*

☐ I have read and understood the information provided in the Participant Information Sheet.

☐ I acknowledge that I can withdraw from this study at any point in time by informing the researchers on site or the Principle Investigator and all my data will be discarded.

☐ My signature below is my acknowledgement that I have agreed to participate in this study according to the above-mentioned terms.

NAME : ______________________________

SIGNATURE : ______________________________

DATE : ______________________________

On a scale of 1-7, please rate your familiarity with your experiment partner

1 2 3 4 5 6 7

Appendix B: Study 1 SPSS Results

Descriptive Statistics

Between-Subjects Factors

		N
Type	Referral	15
	Note	17
	Others	7
	Review	42
	Trace	71

Dependent Variable: Information Exchanged

Type	Mean	Std. Deviation	N
Referral	9.47	3.642	15
Note	9.12	2.088	17
Others	7.57	4.353	7
Review	11.90	2.721	42
Trace	8.99	2.806	71
Total	9.79	3.153	152

Dependent Variable: Information Voluntarily Delivered

Type	Mean	Std. Deviation	N
Referral	8.93	3.390	15
Note	8.76	1.921	17
Others	7.43	4.276	7
Review	10.90	2.792	42
Trace	8.65	2.737	71
Total	9.26	2.982	152

Dependent Variable: Information Prompted

Type	Mean	Std. Deviation	N
Referral	.73	.799	15
Note	.82	.883	17
Others	.29	.488	7
Review	2.02	1.405	42
Trace	1.00	1.134	71
Total	1.20	1.252	152

Dependent Variable: Information Written Down

Type	Mean	Std. Deviation	N
Referral	5.80	1.989	10
Note	5.27	2.240	11
Others	3.17	1.472	6
Review	6.64	2.181	28
Trace	5.15	1.888	47
Total	5.52	2.147	102

Dependent Variable: Handover Quality

Type	Mean	Std. Deviation	N
Referral	3.88	.991	8
Note	4.00	.000	9
Others	4.33	.577	3
Review	3.67	.777	33
Trace	3.44	.867	45
Total	3.63	.817	98

Dependent Variable:Duration

Type	Mean	Std. Deviation	N
Referral	75.20	62.045	15
Note	77.47	53.094	17
Others	50.43	26.913	7
Review	106.38	52.471	42
Trace	67.82	43.084	71
Total	79.48	50.981	152

Ranks

typenumber		N	Mean Rank
Rate	Referral	8	58.06
	Note	9	62.00
	Others	4	70.00
	Review	32	49.88
	Trace	45	43.39
	Total	98	
Duration	Referral	15	66.47
	Note	17	72.32
	Others	8	51.38
	Review	41	105.05
	Trace	71	65.96
	Total	152	
Discussed	Referral	15	72.53
	Note	17	66.06
	Others	8	62.56
	Review	41	106.15
	Trace	71	64.29
	Total	152	
Delivered	Referral	15	72.43
	Note	17	67.94
	Others	8	55.00
	Review	41	102.41
	Trace	71	66.87
	Total	152	
Prompted	Referral	15	59.63
	Note	17	66.15
	Others	8	56.81
	Review	41	101.79
	Trace	71	70.15
	Total	152	
Written	Referral	10	55.10
	Note	11	50.00
	Others	7	24.00
	Review	27	68.06
	Trace	47	45.67
	Total	102	

Test Statistics[a,b]

	Rate	Duration	Discussed	Delivered	Prompted	Written
Chi-square	8.213	24.853	26.212	20.541	21.575	17.019
df	4	4	4	4	4	4
Asymp. Sig.	.084	.000	.000	.000	.000	.002

a. Kruskal Wallis Test

b. Grouping Variable: typenumber

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Rate is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.087	Retain the null hypothesis.
2	The distribution of Duration is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of Discussed is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
4	The distribution of Delivered is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis.
5	The distribution of Prompted is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
6	The distribution of Written is the same across categories of Type.	Independent-Samples Kruskal-Wallis Test	.002	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Rate – handover quality, Duration – handover duration, Discussed – Information Exchanged, Delivered – Information Voluntarily Delivered, Prompted – Information Prompted, Written – Information Written Down.

Appendix C: Study 2 SPSS Results

Effect of IAC (Low, mid, high) and picture type (AA/BC') on hit rate

Within-Subjects Factors

Measure: hit rate

AABC	LMH	Dependent Variable
1	1	AAcorrectL
	2	AAcorrectM
	3	AAcorrectH
2	1	BCcorrectL
	2	BCcorrectM
	3	BCcorrectH

Descriptive Statistics

	Mean	Std. Deviation	N
AAcorrectL	7.9167	1.48083	36
AAcorrectM	8.0278	1.36248	36
AAcorrectH	8.2778	1.44640	36
BCcorrectL	7.5000	1.81265	36
BCcorrectM	7.4444	1.57561	36
BCcorrectH	7.7222	1.64943	36

Tests of Within-Subjects Effects

Measure: Correct rate

Source		F	Sig.
AABC	Sphericity Assumed	7.244	.011
	Greenhouse-Geisser	7.244	.011
	Huynh-Feldt	7.244	.011
	Lower-bound	7.244	.011
LMH	Sphericity Assumed	1.120	.332
	Greenhouse-Geisser	1.120	.331
	Huynh-Feldt	1.120	.332
	Lower-bound	1.120	.297
AABC * LMH	Sphericity Assumed	.073	.930
	Greenhouse-Geisser	.073	.925
	Huynh-Feldt	.073	.930
	Lower-bound	.073	.789

Effect of IAC (Low, mid, high) and picture type (AA/BC') on confidence judgment

Within-Subjects Factors

Measure: confidence

AABC	LMH	Dependent Variable
1	1	AAconsoL
	2	AAconsoM
	3	AAconsoH
2	1	BCconsoL
	2	BCconsoM
	3	BCconsoH

Descriptive Statistics

	Mean	Std. Deviation	N
AAconsoL	78.8750	9.98383	36
AAconsoM	78.9089	8.70042	36
AAconsoH	79.0833	10.07224	36
BCconsoL	78.3981	9.98434	36
BCconsoM	79.6597	8.55051	36
BCconsoH	78.7500	9.76693	36

Tests of Within-Subjects Effects

Measure: confidence

Source		F	Sig.
AABC	Sphericity Assumed	.001	.982
	Greenhouse-Geisser	.001	.982
	Huynh-Feldt	.001	.982
	Lower-bound	.001	.982
LMH	Sphericity Assumed	.164	.849
	Greenhouse-Geisser	.164	.825
	Huynh-Feldt	.164	.835
	Lower-bound	.164	.688
AABC * LMH	Sphericity Assumed	.192	.826
	Greenhouse-Geisser	.192	.821
	Huynh-Feldt	.192	.826
	Lower-bound	.192	.664

Effect of IAC (Low, mid, high) and picture type (AA/BC') on overconfidence

Within-Subjects Factors

Measure: overconfidence

AABC	LMH	Dependent Variable
1	1	AAmetaL
	2	AAmetaM
	3	AAmetaH
2	1	BCmetaL
	2	BCmetaM
	3	BCmetaH

Descriptive Statistics

	Mean	Std. Deviation	N
AAmetaL	-.2917	17.05553	36
AAmetaM	-1.3192	11.67228	36
AAmetaH	-3.6944	10.62294	36
BCmetaL	3.3981	15.93356	36
BCmetaM	5.2153	17.07872	36
BCmetaH	1.5278	14.21365	36

Tests of Within-Subjects Effects

Measure: overconfidence

Source		F	Sig.
AABC	Sphericity Assumed	7.501	.010
	Greenhouse-Geisser	7.501	.010
	Huynh-Feldt	7.501	.010
	Lower-bound	7.501	.010
LMH	Sphericity Assumed	1.357	.264
	Greenhouse-Geisser	1.357	.264
	Huynh-Feldt	1.357	.264
	Lower-bound	1.357	.252
AABC * LMH	Sphericity Assumed	.198	.821
	Greenhouse-Geisser	.198	.797
	Huynh-Feldt	.198	.808
	Lower-bound	.198	.659

Effect of trial sequence (1st, 2nd, 3rd) and picture type (AA/BC') on hit rate

Within-Subjects Factors

Measure: hit rate

AABC	Trialindex	Dependent Variable
1	1	AAcorrect1
	2	AAcorrect2
	3	AAcorrect3
2	1	BCcorrect1
	2	BCcorrect2
	3	BCcorrect3

Descriptive Statistics

	Mean	Std. Deviation	N
AAcorrect1	7.9722	1.38329	36
AAcorrect2	8.2222	1.28976	36
AAcorrect3	8.0278	1.61221	36
BCcorrect1	7.1389	1.53349	36
BCcorrect2	7.5556	1.69780	36
BCcorrect3	7.9722	1.71524	36

Tests of Within-Subjects Effects

Measure: hit rate

Source		F	Sig.
AABC	Sphericity Assumed	7.244	.011
	Greenhouse-Geisser	7.244	.011
	Huynh-Feldt	7.244	.011
	Lower-bound	7.244	.011
Trialindex	Sphericity Assumed	2.394	.099
	Greenhouse-Geisser	2.394	.100
	Huynh-Feldt	2.394	.099
	Lower-bound	2.394	.131
AABC * Trialindex	Sphericity Assumed	1.597	.210
	Greenhouse-Geisser	1.597	.214
	Huynh-Feldt	1.597	.213
	Lower-bound	1.597	.215

Simple effect analysis of picture type in each level of trialindex

Measure: hit rate

Trialindex	(I) AABC	(J) AABC	Mean Difference (I-J)	Std. Error	Sig.[a]
1	1	2	.833*	.297	.008
	2	1	-.833*	.297	.008
2	1	2	.667*	.279	.022
	2	1	-.667*	.279	.022
3	1	2	.056	.394	.889
	2	1	-.056	.394	.889

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Effect of trial sequence (1st, 2nd, 3rd) and picture type (AA/BC') on confidence

Within-Subjects Factors

Measure: confidence

AABC	Trialindex	Dependent Variable
1	1	AAconso1
	2	AAconso2
	3	AAconso3
2	1	BCconso1
	2	BCconso2
	3	BCconso3

Descriptive Statistics

	Mean	Std. Deviation	N
AAconso1	75.0444	8.53703	36
AAconso2	79.9444	8.77967	36
AAconso3	81.9283	9.96094	36
BCconso1	78.7869	8.67584	36
BCconso2	77.6389	9.62483	36
BCconso3	80.3819	9.86163	36

Tests of Within-Subjects Effects

Measure: confidence

Source		F	Sig.
AABC	Sphericity Assumed	.002	.967
	Greenhouse-Geisser	.002	.967
	Huynh-Feldt	.002	.967
	Lower-bound	.002	.967
Trialindex	Sphericity Assumed	8.661	.000
	Greenhouse-Geisser	8.661	.001
	Huynh-Feldt	8.661	.001
	Lower-bound	8.661	.006
AABC * Trialindex	Sphericity Assumed	5.353	.007
	Greenhouse-Geisser	5.353	.010
	Huynh-Feldt	5.353	.009
	Lower-bound	5.353	.027

Simple effect analysis of picture type in each level of trialindex

Measure: confidence

Trialindex	(I) AABC	(J) AABC	Mean Difference (I-J)	Std. Error	Sig.[a]
1	1	2	-3.742*	1.451	.014
	2	1	3.742*	1.451	.014
2	1	2	2.306	1.494	.132
	2	1	-2.306	1.494	.132
3	1	2	1.546	1.401	.277
	2	1	-1.546	1.401	.277

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Effect of trial sequence (1st, 2nd, 3rd) and picture type (AA/BC') on overconfidence

Within-Subjects Factors

Measure: overconfidence

AABC	Trialindex	Dependent Variable
1	1	AAmeta1
	2	AAmeta2
	3	AAmeta3
2	1	BCmeta1
	2	BCmeta2
	3	BCmeta3

Descriptive Statistics

	Mean	Std. Deviation	N
AAmeta1	-4.6778	12.96056	36
AAmeta2	-2.2778	11.25195	36
AAmeta3	1.6506	15.18806	36
BCmeta1	7.3981	15.21006	36
BCmeta2	2.0833	17.07693	36
BCmeta3	.6597	14.34114	36

Tests of Within-Subjects Effects

Measure: overconfidence

Source		F	Sig.
AABC	Sphericity Assumed	7.501	.010
	Greenhouse-Geisser	7.501	.010
	Huynh-Feldt	7.501	.010
	Lower-bound	7.501	.010
Trialindex	Sphericity Assumed	.302	.740
	Greenhouse-Geisser	.302	.738
	Huynh-Feldt	.302	.740
	Lower-bound	.302	.586
AABC * Trialindex	Sphericity Assumed	4.750	.012
	Greenhouse-Geisser	4.750	.015
	Huynh-Feldt	4.750	.013
	Lower-bound	4.750	.036

Simple effect analysis of picture type in each level of trialindex

Measure: overconfidence

Trialindex	(I) AABC	(J) AABC	Mean Difference (I-J)	Std. Error	Sig.[a]
1	1	2	-12.076*	2.739	.000
	2	1	12.076*	2.739	.000
2	1	2	-4.361	2.633	.107
	2	1	4.361	2.633	.107
3	1	2	.991	3.786	.795
	2	1	-.991	3.786	.795

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Effect of Trialindex (1st, 2nd, 3rd) and picture type (AA/BC') on number of access

Within-Subjects Factors

Measure: Number of access

AABC	Trialindex	Dependent Variable
1	1	AAaccess1
	2	AAaccess2
	3	AAaccess3
2	1	BCaccess1
	2	BCaccess2
	3	BCaccess3

Descriptive Statistics

	Mean	Std. Deviation	N
AAaccess1	6.1944	2.37630	36
AAaccess2	6.0278	2.44349	36
AAaccess3	5.3889	2.80079	36
BCaccess1	5.1389	2.71664	36
BCaccess2	5.2778	3.03890	36
BCaccess3	5.3889	2.96915	36

Tests of Within-Subjects Effects

Measure: Number of access

Source		F	Sig.
AABC	Sphericity Assumed	4.467	.042
	Greenhouse-Geisser	4.467	.042
	Huynh-Feldt	4.467	.042
	Lower-bound	4.467	.042
Trialindex	Sphericity Assumed	.395	.675
	Greenhouse-Geisser	.395	.650
	Huynh-Feldt	.395	.660
	Lower-bound	.395	.534
AABC * Trialindex	Sphericity Assumed	1.429	.246
	Greenhouse-Geisser	1.429	.247
	Huynh-Feldt	1.429	.246
	Lower-bound	1.429	.240

Simple effect analysis of picture type in each level of trialindex

Measure: Number of access

Trialindex	(I) AABC	(J) AABC	Mean Difference (I-J)	Std. Error	Sig.[a]
1	1	2	1.056*	.471	.032
	2	1	-1.056*	.471	.032
2	1	2	.750	.479	.126
	2	1	-.750	.479	.126
3	1	2	.000	.453	1.000
	2	1	.000	.453	1.000

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Effect of Trialindex (1st, 2nd, 3rd) and picture type (AA/BC') on test 2 recognition performance

Within-Subjects Factors

Measure:Result

AABC	Trialindex	Dependent Variable
1	1	AAresult1
	2	AAresult2
	3	AAresult3
2	1	BCresult1
	2	BCresult2
	3	Bcresult3

Descriptive Statistics

	Mean	Std. Deviation	N
AAresult1	9.1111	1.18990	36
AAresult2	9.0278	1.36248	36
AAresult3	9.2500	.90633	36
BCresult1	8.3333	1.54919	36
BCresult2	8.7500	1.61024	36
Bcresult3	8.9722	1.15847	36

Tests of Within-Subjects Effects

Measure: Result

Source		F	Sig.
AABC	Sphericity Assumed	8.960	.005
	Greenhouse-Geisser	8.960	.005
	Huynh-Feldt	8.960	.005
	Lower-bound	8.960	.005
Trialindex	Sphericity Assumed	2.586	.082
	Greenhouse-Geisser	2.586	.086
	Huynh-Feldt	2.586	.083
	Lower-bound	2.586	.117
AABC * Trialindex	Sphericity Assumed	1.445	.243
	Greenhouse-Geisser	1.445	.243
	Huynh-Feldt	1.445	.243
	Lower-bound	1.445	.237

Simple effect analysis of picture type in each level of trialindex

Measure: Result

Trialindex	(I) AABC	(J) AABC	Mean Difference (I-J)	Std. Error	Sig.[a]
1	1	2	.778*	.259	.005
	2	1	-.778*	.259	.005
2	1	2	.278	.283	.334
	2	1	-.278	.283	.334
3	1	2	.278	.185	.143
	2	1	-.278	.185	.143

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Appendix D: Study 3 SPSS Results

Effect of IAC (Low/high) and accountability (Low/high) on hit rate

Descriptive Statistics

	Mean	Std. Deviation	N
accuracyLL	54.6667	11.05888	30
accuracyLH	55.6667	11.35124	30
accuracyHL	59.3333	12.84747	30
accuracyHH	63.0000	15.56964	30

Tests of Within-Subjects Effects

Measure: hit rate

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ACCT	Sphericity Assumed	1080.000	1	1080.000	7.695	.010
	Greenhouse-Geisser	1080.000	1.000	1080.000	7.695	.010
	Huynh-Feldt	1080.000	1.000	1080.000	7.695	.010
	Lower-bound	1080.000	1.000	1080.000	7.695	.010
Error(ACCT)	Sphericity Assumed	4070.000	29	140.345		
	Greenhouse-Geisser	4070.000	29.000	140.345		
	Huynh-Feldt	4070.000	29.000	140.345		
	Lower-bound	4070.000	29.000	140.345		
IAC	Sphericity Assumed	163.333	1	163.333	1.011	.323

	Greenhouse-Geisser	163.333	1.000	163.333	1.011	.323
	Huynh-Feldt	163.333	1.000	163.333	1.011	.323
	Lower-bound	163.333	1.000	163.333	1.011	.323
Error(IAC)	Sphericity Assumed	4686.667	29	161.609		
	Greenhouse-Geisser	4686.667	29.000	161.609		
	Huynh-Feldt	4686.667	29.000	161.609		
	Lower-bound	4686.667	29.000	161.609		
ACCT * IAC	Sphericity Assumed	53.333	1	53.333	.287	.596
	Greenhouse-Geisser	53.333	1.000	53.333	.287	.596
	Huynh-Feldt	53.333	1.000	53.333	.287	.596
	Lower-bound	53.333	1.000	53.333	.287	.596
Error(ACCT*IAC)	Sphericity Assumed	5396.667	29	186.092		
	Greenhouse-Geisser	5396.667	29.000	186.092		
	Huynh-Feldt	5396.667	29.000	186.092		
	Lower-bound	5396.667	29.000	186.092		

Effect of IAC (Low/high) and accountability (Low/high) on confidence judgment

Descriptive Statistics

	Mean	Std. Deviation	N
confLL	69.1700	14.52164	30
confLH	68.9283	13.71757	30
confHL	68.2323	16.98712	30
confHH	71.1133	12.77813	30

Tests of Within-Subjects Effects

Measure:confidence

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ACCT	Sphericity Assumed	11.669	1	11.669	.249	.622
	Greenhouse-Geisser	11.669	1.000	11.669	.249	.622
	Huynh-Feldt	11.669	1.000	11.669	.249	.622
	Lower-bound	11.669	1.000	11.669	.249	.622
Error(ACCT)	Sphericity Assumed	1360.118	29	46.901		
	Greenhouse-Geisser	1360.118	29.000	46.901		
	Huynh-Feldt	1360.118	29.000	46.901		
	Lower-bound	1360.118	29.000	46.901		
IAC	Sphericity Assumed	52.246	1	52.246	.726	.401
	Greenhouse-Geisser	52.246	1.000	52.246	.726	.401
	Huynh-Feldt	52.246	1.000	52.246	.726	.401
	Lower-bound	52.246	1.000	52.246	.726	.401
Error(IAC)	Sphericity Assumed	2086.289	29	71.941		

	Greenhouse-Geisser	2086.289	29.000	71.941		
	Huynh-Feldt	2086.289	29.000	71.941		
	Lower-bound	2086.289	29.000	71.941		
ACCT * IAC	Sphericity Assumed	73.133	1	73.133	.644	.429
	Greenhouse-Geisser	73.133	1.000	73.133	.644	.429
	Huynh-Feldt	73.133	1.000	73.133	.644	.429
	Lower-bound	73.133	1.000	73.133	.644	.429
Error(ACCT*IAC)	Sphericity Assumed	3294.468	29	113.602		
	Greenhouse-Geisser	3294.468	29.000	113.602		
	Huynh-Feldt	3294.468	29.000	113.602		
	Lower-bound	3294.468	29.000	113.602		

Effect of IAC (Low/high) and accountability (Low/high) on overconfidence

Descriptive Statistics

	Mean	Std. Deviation	N
OCLL	14.5033	15.35750	30
OCLH	13.2617	16.79725	30
OCHL	8.8990	19.17944	30
OCHH	8.1133	16.72448	30

Tests of Within-Subjects Effects

Measure:overconfidence

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ACCT	Sphericity Assumed	867.149	1	867.149	4.395	.045
	Greenhouse-Geisser	867.149	1.000	867.149	4.395	.045
	Huynh-Feldt	867.149	1.000	867.149	4.395	.045
	Lower-bound	867.149	1.000	867.149	4.395	.045
Error(ACCT)	Sphericity Assumed	5722.338	29	197.322		
	Greenhouse-Geisser	5722.338	29.000	197.322		
	Huynh-Feldt	5722.338	29.000	197.322		
	Lower-bound	5722.338	29.000	197.322		
IAC	Sphericity Assumed	30.826	1	30.826	.168	.685
	Greenhouse-Geisser	30.826	1.000	30.826	.168	.685
	Huynh-Feldt	30.826	1.000	30.826	.168	.685
	Lower-bound	30.826	1.000	30.826	.168	.685
Error(IAC)	Sphericity Assumed	5320.309	29	183.459		

	Greenhouse-Geisser	5320.309	29.000	183.459		
	Huynh-Feldt	5320.309	29.000	183.459		
	Lower-bound	5320.309	29.000	183.459		
ACCT * IAC	Sphericity Assumed	1.560	1	1.560	.007	.935
	Greenhouse-Geisser	1.560	1.000	1.560	.007	.935
	Huynh-Feldt	1.560	1.000	1.560	.007	.935
	Lower-bound	1.560	1.000	1.560	.007	.935
Error(ACCT*IAC)	Sphericity Assumed	6666.942	29	229.895		
	Greenhouse-Geisser	6666.942	29.000	229.895		
	Huynh-Feldt	6666.942	29.000	229.895		
	Lower-bound	6666.942	29.000	229.895		

Effect of IAC (Low/high) and accountability (Low/high) on number of access attempts

Descriptive Statistics

	Mean	Std. Deviation	N
IAALL	6.3214	3.62147	28
IAALH	1.4643	2.02726	28
IAAHL	6.0357	3.40460	28
IAAHH	2.1071	2.28261	28

Tests of Within-Subjects Effects

Measure: information access attempts

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ACCT	Sphericity Assumed	.893	1	.893	.209	.652
	Greenhouse-Geisser	.893	1.000	.893	.209	.652
	Huynh-Feldt	.893	1.000	.893	.209	.652
	Lower-bound	.893	1.000	.893	.209	.652
Error(ACCT)	Sphericity Assumed	115.607	27	4.282		
	Greenhouse-Geisser	115.607	27.000	4.282		
	Huynh-Feldt	115.607	27.000	4.282		
	Lower-bound	115.607	27.000	4.282		
IAC	Sphericity Assumed	540.321	1	540.321	97.793	.000
	Greenhouse-Geisser	540.321	1.000	540.321	97.793	.000
	Huynh-Feldt	540.321	1.000	540.321	97.793	.000
	Lower-bound	540.321	1.000	540.321	97.793	.000
Error(IAC)	Sphericity Assumed	149.179	27	5.525		

	Greenhouse-Geisser	149.179	27.000	5.525		
	Huynh-Feldt	149.179	27.000	5.525		
	Lower-bound	149.179	27.000	5.525		
ACCT * IAC	Sphericity Assumed	6.036	1	6.036	2.313	.140
	Greenhouse-Geisser	6.036	1.000	6.036	2.313	.140
	Huynh-Feldt	6.036	1.000	6.036	2.313	.140
	Lower-bound	6.036	1.000	6.036	2.313	.140
Error(ACCT*IAC)	Sphericity Assumed	70.464	27	2.610		
	Greenhouse-Geisser	70.464	27.000	2.610		
	Huynh-Feldt	70.464	27.000	2.610		
	Lower-bound	70.464	27.000	2.610		

Effect of IAC (Low/high) and accountability (Low/high) on test 2 performance

Descriptive Statistics

	Mean	Std. Deviation	N
perfLL	78.6667	19.95397	30
perfLH	62.3333	18.69600	30
perfHL	78.3333	15.99210	30
perfHH	69.0000	18.07074	30

Tests of Within-Subjects Effects

Measure: test 2 performance

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ACCT	Sphericity Assumed	300.833	1	300.833	2.706	.111
	Greenhouse-Geisser	300.833	1.000	300.833	2.706	.111
	Huynh-Feldt	300.833	1.000	300.833	2.706	.111
	Lower-bound	300.833	1.000	300.833	2.706	.111
Error(ACCT)	Sphericity Assumed	3224.167	29	111.178		
	Greenhouse-Geisser	3224.167	29.000	111.178		
	Huynh-Feldt	3224.167	29.000	111.178		
	Lower-bound	3224.167	29.000	111.178		
IAC	Sphericity Assumed	4940.833	1	4940.833	20.226	.000
	Greenhouse-Geisser	4940.833	1.000	4940.833	20.226	.000
	Huynh-Feldt	4940.833	1.000	4940.833	20.226	.000
	Lower-bound	4940.833	1.000	4940.833	20.226	.000
Error(IAC)	Sphericity Assumed	7084.167	29	244.282		

	Greenhouse-Geisser	7084.167	29.000	244.282		
	Huynh-Feldt	7084.167	29.000	244.282		
	Lower-bound	7084.167	29.000	244.282		
ACCT * IAC	Sphericity Assumed	367.500	1	367.500	2.391	.133
	Greenhouse-Geisser	367.500	1.000	367.500	2.391	.133
	Huynh-Feldt	367.500	1.000	367.500	2.391	.133
	Lower-bound	367.500	1.000	367.500	2.391	.133
Error(ACCT*IAC)	Sphericity Assumed	4457.500	29	153.707		
	Greenhouse-Geisser	4457.500	29.000	153.707		
	Huynh-Feldt	4457.500	29.000	153.707		
	Lower-bound	4457.500	29.000	153.707		

Appendix E: Study 4 SPSS Results

Effect of reviewability on memory retention

Descriptive Statistics

	complexity	Mean	Std. Deviation	N
memory_lowreviewability	Complex	18.3667	4.32063	12
	Simple	21.4500	4.87153	12
	Total	19.9083	4.77055	24
memory_highreviewability	Complex	22.1667	2.96995	12
	Simple	23.4000	4.13851	12
	Total	22.7833	3.57864	24

Tests of Between-Subjects Effects

Measure: memory retention

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	21870.941	1	21870.941	923.571	.000
complexity	55.901	1	55.901	2.361	.139
Error	520.978	22	23.681		

Tests of Within-Subjects Effects

Measure: memory retention

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Reviewability	Sphericity Assumed	99.188	1	99.188	9.453	.006
	Greenhouse-Geisser	99.188	1.000	99.188	9.453	.006
	Huynh-Feldt	99.188	1.000	99.188	9.453	.006
	Lower-bound	99.188	1.000	99.188	9.453	.006
Reviewability * complexity	Sphericity Assumed	10.267	1	10.267	.979	.333
	Greenhouse-Geisser	10.267	1.000	10.267	.979	.333
	Huynh-Feldt	10.267	1.000	10.267	.979	.333
	Lower-bound	10.267	1.000	10.267	.979	.333
Error(Reviewability)	Sphericity Assumed	230.845	22	10.493		
	Greenhouse-Geisser	230.845	22.000	10.493		
	Huynh-Feldt	230.845	22.000	10.493		
	Lower-bound	230.845	22.000	10.493		

Effect of reviewability on case comprehension

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
comprehension_lowreviewability		Complex	6.8333	1.74946	12
		Simple	7.0833	2.35327	12
		Total	6.9583	2.03190	24
comprehension_highreviewability		Complex	7.0000	2.37410	12
		Simple	7.0833	2.77843	12
		Total	7.0417	2.52774	24

Tests of Between-Subjects Effects

Measure: case comprehension

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	2352.000	1	2352.000	318.098	.000
complexity	.333	1	.333	.045	.834
Error	162.667	22	7.394		

Tests of Within-Subjects Effects

Measure: case comprehension

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Reviewability	Sphericity Assumed	.083	1	.083	.023	.880
	Greenhouse-Geisser	.083	1.000	.083	.023	.880
	Huynh-Feldt	.083	1.000	.083	.023	.880
	Lower-bound	.083	1.000	.083	.023	.880
Reviewability * complexity	Sphericity Assumed	.083	1	.083	.023	.880
	Greenhouse-Geisser	.083	1.000	.083	.023	.880
	Huynh-Feldt	.083	1.000	.083	.023	.880
	Lower-bound	.083	1.000	.083	.023	.880
Error(Reviewability)	Sphericity Assumed	78.833	22	3.583		
	Greenhouse-Geisser	78.833	22.000	3.583		
	Huynh-Feldt	78.833	22.000	3.583		
	Lower-bound	78.833	22.000	3.583		

Effect of reviewability on subjective evaluation

Descriptive Statistics

	complexity		Mean	Std. Deviation	N
subjective_lowreviewability		Complex	4.6667	1.15470	12
		Simple	5.3333	1.23091	12
		Total	5.0000	1.21584	24
subjective_highreviewability		Complex	5.3750	1.06867	12
		Simple	5.7500	.75378	12
		Total	5.5625	.92446	24

Tests of Between-Subjects Effects

Measure: subject evaluation

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	1338.797	1	1338.797	947.884	.000
complexity	3.255	1	3.255	2.305	.143
Error	31.073	22	1.412		

Tests of Within-Subjects Effects

Measure: subject evaluation

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Reviewability	Sphericity Assumed	3.797	1	3.797	4.380	.048
	Greenhouse-Geisser	3.797	1.000	3.797	4.380	.048
	Huynh-Feldt	3.797	1.000	3.797	4.380	.048
	Lower-bound	3.797	1.000	3.797	4.380	.048
Reviewability * complexity	Sphericity Assumed	.255	1	.255	.294	.593
	Greenhouse-Geisser	.255	1.000	.255	.294	.593
	Huynh-Feldt	.255	1.000	.255	.294	.593
	Lower-bound	.255	1.000	.255	.294	.593
Error(Reviewability)	Sphericity Assumed	19.073	22	.867		
	Greenhouse-Geisser	19.073	22.000	.867		
	Huynh-Feldt	19.073	22.000	.867		
	Lower-bound	19.073	22.000	.867		

Effect of interactivity on memory retention

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
memory_lowinteractivity		Complex	14.7000	7.99932	12
		Simple	18.2500	5.37883	12
		Total	16.4750	6.90855	24
memory_highinteractivity		Complex	20.5833	5.99179	12
		Simple	20.2667	4.31010	12
		Total	20.4250	5.10696	24

Tests of Between-Subjects Effects

Measure: memory retention

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	16339.320	1	16339.320	363.171	.000
complexity	31.363	1	31.363	.697	.413
Error	989.797	22	44.991		

Tests of Within-Subjects Effects

Measure: memory retention

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Interactivity	Sphericity Assumed	187.230	1	187.230	6.522	.018
	Greenhouse-Geisser	187.230	1.000	187.230	6.522	.018
	Huynh-Feldt	187.230	1.000	187.230	6.522	.018
	Lower-bound	187.230	1.000	187.230	6.522	.018
Interactivity * complexity	Sphericity Assumed	44.853	1	44.853	1.562	.224
	Greenhouse-Geisser	44.853	1.000	44.853	1.562	.224
	Huynh-Feldt	44.853	1.000	44.853	1.562	.224
	Lower-bound	44.853	1.000	44.853	1.562	.224
Error(Interactivity)	Sphericity Assumed	631.597	22	28.709		
	Greenhouse-Geisser	631.597	22.000	28.709		
	Huynh-Feldt	631.597	22.000	28.709		
	Lower-bound	631.597	22.000	28.709		

Effect of interactivity on case comprehension

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
comprehension_lowinteractivity		Complex	5.1667	2.08167	12
		Simple	5.6667	2.80692	12
		Total	5.4167	2.43018	24
comprehension_highinteractivity		Complex	6.7500	1.76455	12
		Simple	6.9167	2.84312	12
		Total	6.8333	2.31567	24

Tests of Between-Subjects Effects

Measure: case comprehension

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	1800.750	1	1800.750	279.153	.000
complexity	1.333	1	1.333	.207	.654
Error	141.917	22	6.451		

Tests of Within-Subjects Effects

Measure: case comprehension

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Interactivity	Sphericity Assumed	24.083	1	24.083	4.584	.044
	Greenhouse-Geisser	24.083	1.000	24.083	4.584	.044
	Huynh-Feldt	24.083	1.000	24.083	4.584	.044
	Lower-bound	24.083	1.000	24.083	4.584	.044
Interactivity * complexity	Sphericity Assumed	.333	1	.333	.063	.803
	Greenhouse-Geisser	.333	1.000	.333	.063	.803
	Huynh-Feldt	.333	1.000	.333	.063	.803
	Lower-bound	.333	1.000	.333	.063	.803
Error(Interactivity)	Sphericity Assumed	115.583	22	5.254		
	Greenhouse-Geisser	115.583	22.000	5.254		
	Huynh-Feldt	115.583	22.000	5.254		
	Lower-bound	115.583	22.000	5.254		

Effect of interactivity on subjective evaluation

Descriptive Statistics

	complexity		Mean	Std. Deviation	N
subjective_lowinteractivity		Complex	5.3333	1.28511	12
		Simple	5.1250	1.68044	12
		Total	5.2292	1.46687	24
subjective_highinteractivity		Complex	5.5833	1.08362	12
		Simple	5.8750	.95644	12
		Total	5.7292	1.01059	24

Tests of Between-Subjects Effects

Measure: subject evaluation

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	1441.021	1	1441.021	774.017	.000
complexity	.021	1	.021	.011	.917
Error	40.958	22	1.862		

Tests of Within-Subjects Effects

Measure: subject evaluation

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Interactivity	Sphericity Assumed	3.000	1	3.000	2.112	.160
	Greenhouse-Geisser	3.000	1.000	3.000	2.112	.160
	Huynh-Feldt	3.000	1.000	3.000	2.112	.160
	Lower-bound	3.000	1.000	3.000	2.112	.160
Interactivity * complexity	Sphericity Assumed	.750	1	.750	.528	.475
	Greenhouse-Geisser	.750	1.000	.750	.528	.475
	Huynh-Feldt	.750	1.000	.750	.528	.475
	Lower-bound	.750	1.000	.750	.528	.475
Error(Interactivity)	Sphericity Assumed	31.250	22	1.420		
	Greenhouse-Geisser	31.250	22.000	1.420		
	Huynh-Feldt	31.250	22.000	1.420		
	Lower-bound	31.250	22.000	1.420		

Effect of visual access to object of reference on memory retention

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
memory_noaccess		Complex	20.5833	5.99179	12
		Simple	20.2667	4.31010	12
		Total	20.4250	5.10696	24
memory_yesaccess		Complex	23.1167	4.09630	12
		Simple	22.9833	4.62932	12
		Total	23.0500	4.27541	24

Tests of Between-Subjects Effects

Measure: memory retention

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	22680.907	1	22680.907	1172.402	.000
complexity	.608	1	.608	.031	.861
Error	425.605	22	19.346		

Tests of Within-Subjects Effects

Measure: memory retention

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Visualaccess	Sphericity Assumed	82.687	1	82.687	3.063	.094
	Greenhouse-Geisser	82.687	1.000	82.687	3.063	.094
	Huynh-Feldt	82.687	1.000	82.687	3.063	.094
	Lower-bound	82.687	1.000	82.687	3.063	.094
Visualaccess * complexity	Sphericity Assumed	.101	1	.101	.004	.952
	Greenhouse-Geisser	.101	1.000	.101	.004	.952
	Huynh-Feldt	.101	1.000	.101	.004	.952
	Lower-bound	.101	1.000	.101	.004	.952
Error(Visualaccess)	Sphericity Assumed	593.972	22	26.999		
	Greenhouse-Geisser	593.972	22.000	26.999		
	Huynh-Feldt	593.972	22.000	26.999		
	Lower-bound	593.972	22.000	26.999		

Effect of visual access to object of reference on case comprehension

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
comprehension_noaccess		Complex	6.7500	1.76455	12
		Simple	6.9167	2.84312	12
		Total	6.8333	2.31567	24
comprehension_yesaccess		Complex	7.3333	2.05971	12
		Simple	7.5000	1.16775	12
		Total	7.4167	1.63964	24

Tests of Between-Subjects Effects

Measure: case comprehension

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	2436.750	1	2436.750	436.137	.000
complexity	.333	1	.333	.060	.809
Error	122.917	22	5.587		

Tests of Within-Subjects Effects

Measure: case comprehension

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Visualaccess	Sphericity Assumed	4.083	1	4.083	1.451	.241
	Greenhouse-Geisser	4.083	1.000	4.083	1.451	.241
	Huynh-Feldt	4.083	1.000	4.083	1.451	.241
	Lower-bound	4.083	1.000	4.083	1.451	.241
Visualaccess * complexity	Sphericity Assumed	.000	1	.000	.000	1.000
	Greenhouse-Geisser	.000	1.000	.000	.000	1.000
	Huynh-Feldt	.000	1.000	.000	.000	1.000
	Lower-bound	.000	1.000	.000	.000	1.000
Error(Visualaccess)	Sphericity Assumed	61.917	22	2.814		
	Greenhouse-Geisser	61.917	22.000	2.814		
	Huynh-Feldt	61.917	22.000	2.814		
	Lower-bound	61.917	22.000	2.814		

Effect of visual access to object of reference on subjective evaluation

Descriptive Statistics

	complexity	Mean	Std. Deviation	N
subject_noaccess	Complex	5.5833	1.08362	12
	Simple	5.8750	.95644	12
	Total	5.7292	1.01059	24
subject_yesaccess	Complex	5.3750	1.06867	12
	Simple	5.7500	.75378	12
	Total	5.5625	.92446	24

Tests of Between-Subjects Effects

Measure: subject evaluation

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	1530.021	1	1530.021	1573.225	.000
complexity	1.333	1	1.333	1.371	.254
Error	21.396	22	.973		

Tests of Within-Subjects Effects

Measure: subject evaluation

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Visualaccess	Sphericity Assumed	.333	1	.333	.360	.555
	Greenhouse-Geisser	.333	1.000	.333	.360	.555
	Huynh-Feldt	.333	1.000	.333	.360	.555
	Lower-bound	.333	1.000	.333	.360	.555
Visualaccess * complexity	Sphericity Assumed	.021	1	.021	.022	.882
	Greenhouse-Geisser	.021	1.000	.021	.022	.882
	Huynh-Feldt	.021	1.000	.021	.022	.882
	Lower-bound	.021	1.000	.021	.022	.882
Error(Visualaccess)	Sphericity Assumed	20.396	22	.927		
	Greenhouse-Geisser	20.396	22.000	.927		
	Huynh-Feldt	20.396	22.000	.927		
	Lower-bound	20.396	22.000	.927		

Effect of visibility on memory retention

Descriptive Statistics					
	complexity		Mean	Std. Deviation	N
memory_novisibility		Complex	20.5833	5.99179	12
		Simple	20.2667	4.31010	12
		Total	20.4250	5.10696	24
memory_yesvisibility		Complex	20.2500	5.94100	12
		Simple	21.2250	5.77520	12
		Total	20.7375	5.75151	24

Tests of Between-Subjects Effects

Measure: memory retention

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
	20332.217	1	20332.217	445.443	.000	445.443	1.000
complexity	1.300	1	1.300	.028	.868	.028	.053
Error	1004.188	22	45.645				

a. Computed using alpha = .05

Tests of Within-Subjects Effects

Measure: memory retention

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Visibility	Sphericity Assumed	1.172	1	1.172	.074	.789	.074	.058
	Greenhouse-Geisser	1.172	1.000	1.172	.074	.789	.074	.058
	Huynh-Feldt	1.172	1.000	1.172	.074	.789	.074	.058
	Lower-bound	1.172	1.000	1.172	.074	.789	.074	.058
Visibility * complexity	Sphericity Assumed	5.005	1	5.005	.314	.581	.314	.084
	Greenhouse-Geisser	5.005	1.000	5.005	.314	.581	.314	.084
	Huynh-Feldt	5.005	1.000	5.005	.314	.581	.314	.084
	Lower-bound	5.005	1.000	5.005	.314	.581	.314	.084
Error(Visibility)	Sphericity Assumed	350.208	22	15.919				
	Greenhouse-Geisser	350.208	22.000	15.919				
	Huynh-Feldt	350.208	22.000	15.919				
	Lower-bound	350.208	22.000	15.919				

a. Computed using alpha = .05

Effect of visibility on case comprehension

Descriptive Statistics

	complexity		Mean	Std. Deviation	N
comprehension_novisibility		Complex	6.7500	1.76455	12
		Simple	6.9167	2.84312	12
		Total	6.8333	2.31567	24
comprehensin_yesvisibility		Complex	6.6667	2.34844	12
		Simple	6.9167	2.99874	12
		Total	6.7917	2.63718	24

Tests of Between-Subjects Effects

Measure: case comprehension

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
	2227.687	1	2227.687	301.982	.000	301.982	1.000
complexity	.521	1	.521	.071	.793	.071	.057
Error	162.292	22	7.377				

a. Computed using alpha = .05

Tests of Within-Subjects Effects

Measure: case comprehension

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Visibility	Sphericity Assumed	.021	1	.021	.004	.951	.004	.050
	Greenhouse-Geisser	.021	1.000	.021	.004	.951	.004	.050
	Huynh-Feldt	.021	1.000	.021	.004	.951	.004	.050
	Lower-bound	.021	1.000	.021	.004	.951	.004	.050
Visibility * complexity	Sphericity Assumed	.021	1	.021	.004	.951	.004	.050
	Greenhouse-Geisser	.021	1.000	.021	.004	.951	.004	.050
	Huynh-Feldt	.021	1.000	.021	.004	.951	.004	.050
	Lower-bound	.021	1.000	.021	.004	.951	.004	.050
Error(Visibility)	Sphericity Assumed	120.458	22	5.475				
	Greenhouse-Geisser	120.458	22.000	5.475				
	Huynh-Feldt	120.458	22.000	5.475				
	Lower-bound	120.458	22.000	5.475				

a. Computed using alpha = .05

Effect of visibility on subjective evaluation

Descriptive Statistics

	complexity	Mean	Std. Deviation	N
subject_novisibility	Complex	5.5833	1.08362	12
	Simple	5.8750	.95644	12
	Total	5.7292	1.01059	24
subject_yesvisibility	Complex	5.3333	.65134	12
	Simple	5.6667	1.23091	12
	Total	5.5000	.97802	24

Tests of Between-Subjects Effects

Measure: subject evaluation

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
	1513.130	1	1513.130	1062.764	.000	1062.764	1.000
complexity	1.172	1	1.172	.823	.374	.823	.140
Error	31.323	22	1.424				

a. Computed using alpha = .05

Tests of Within-Subjects Effects

Measure: subject evaluation

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Visibility	Sphericity Assumed	.630	1	.630	1.067	.313	1.067	.167
	Greenhouse-Geisser	.630	1.000	.630	1.067	.313	1.067	.167
	Huynh-Feldt	.630	1.000	.630	1.067	.313	1.067	.167
	Lower-bound	.630	1.000	.630	1.067	.313	1.067	.167
Visibility * complexity	Sphericity Assumed	.005	1	.005	.009	.926	.009	.051
	Greenhouse-Geisser	.005	1.000	.005	.009	.926	.009	.051
	Huynh-Feldt	.005	1.000	.005	.009	.926	.009	.051
	Lower-bound	.005	1.000	.005	.009	.926	.009	.051
Error(Visibility)	Sphericity Assumed	12.990	22	.590				
	Greenhouse-Geisser	12.990	22.000	.590				
	Huynh-Feldt	12.990	22.000	.590				
	Lower-bound	12.990	22.000	.590				

a. Computed using alpha = .05