A user study on the information search behaviour of medical students

Anushia Inthiran[1*], Saadat M. Alhashmi[2] and **Pervaiz K. Ahmed[1]**
[1]School of Information Technology, Monash University Malaysia,
Jalan Lagoon Selatan Bandar Sunway 46150
Petaling Jaya, MALAYSIA.
[2]School of Engineering and Computer Science,
The Abu Dhabi University P.O. Box 59911,
Abu Dhabi, UAE P.O. Box 1790, Al Ain, UNITED ARAB EMIRATES.
e-mail: *anushia.inthiran@monash.edu (corresponding author);
saadat.alhashmi@adu.ac.ae; pervaiz.ahmed@monash.edu

ABSTRACT

This study assesses the information search behaviour of medical students when searching for a medical type task. Previous research provide information on the general search traits of medical students. This study shows how medical students search for medical information online for specific medical tasks. It examines the demographic variables that influence medical students search behaviour when searching across varying levels of task difficulty. Detailed aspects of a search session are analysed to examine this effect. Search sessions are broadly classified into: (a) querying details, (b) search results interaction details and (c) querying versus clicking behaviour and task completion time. An interactive information retrieval experiment type methodology (IIR) is used to study the interactive searching behaviour. A total of sixty (60) medical students participated in this study. Personal task and simulated situations were used to invoke the information search process. Results indicate that medical students issued long queries, were active in locating results and were slow and unproductive when searching on a difficult task. Demographic variables did not influence the search behaviour of medical students when searching across varying levels of task difficulty. Results of this study provide an understanding of the information search behaviour of medical students when searching for medical information online across varying levels of task difficulty. It also shows that demographic variables did not influence search behaviour when searching across varying level of task difficulty.

Keywords: Information searching; Information behaviour; Medical Students; Task Difficulty; Medical Search; User Study.

INTRODUCTION

One of the earliest research investigating medical students search behaviour was conducted to analyse the effects of medical search training amongst medical students (Mitchell, Johnson and Proud 1990; Mitchell et al. 1992). Research conducted in the 1990's found that medical students who were trained to use a system called *Grateful Med*[1] became independent users of the system and used Medical Subject Headings (MeSH)

[1]Grateful Med is a retired medical database containing information from MEDLINE and the National Library of Medicine.

vocabulary terms effectively (Mitchell, Johnson and Proud, 1990; Mitchell et al. 1992). However, a later study indicates training only had a short-term benefit to improving a medical student search session (Shelstad et al. 1994). In a comparative study, Hersh et al. (2002) illustrated the differences in search behaviour of medical students and nurse practitioners. Medical students were able to answer more questions correctly in comparison to nurse practitioners. Medical students and nurse practitioners were able to obtain comparable measurements of precision and recall but neither measurement were associated to the correctness of answers. Other studies focus on medical students' education and training needs (Peterson et al. 2004; Gruppen, Rana and Arndt 2005; Llic, Tepper and Misso 2006; Shurtz 2009). E-learning initiatives and digital textbooks have also paved their way into medical education. Schilling et al. (2006) and Gormley et al. (2009) indicate usage of e-learning technologies and applications have been shown to increase medical students' ability to perform better medical searchers. Abderrahim et al. (2015) indicate medical students undergoing residency have high theoretical learning quality when they frequently referred to and understood clinical guidelines. Thus, it is important for students to use clinical guidelines when searching for medical information online.

While interactive search behaviour of medical students was actively research in the 1990's and early 2000's, no new research studies have emerged since 2008. We believe one reason for this is that research focus in relation to medical students search behaviour has shifted to education and training initiatives. With the exception of a small scale exploratory research study in relation to the effects of task difficulty on medical students querying behaviour (Inthiran, Alhashmi and Ahmed 2011), there was scarce literature in relation to the interactive search behaviour of medical students. Similarly, we were unable to locate information in relation to the influence of demographic variables on medical students' medical information searching behaviour. On the contrary there have been a number of seminal papers in relation to the influence of task difficulty on non-medical type searching (Bystrom 2002; Li and Belkin 2008; Aula, Khan and Guan 2010; Liu et al. 2012). Similarly, Hu, Zeng and Niu (2007), Jones et al. (2007) and Weber and Castillo (2010) found demographic variables influenced non-medical search behaviour. As a result of these studies, we are now able to better understand search behaviour and the background information of the searcher. In return, researchers were able to provide searchers with relevant information retrieval assistance and develop search behaviour profiles based on the categorization of search behaviour.

Medical domains are huge and undergo continuous expansion (Nadkarni 2000; Can and Baykal 2007). Thus, it is important that medical students keep abreast with updated information to improve learning outcomes towards becoming better doctors. In addition, performing a non-medical type search is unique and dissimilar to non-medical searching (Hersh 2009). As such, examining how medical students perform a medical search based on varying levels of task difficulty will provide us with theoretical understanding of medical students search behaviour. It is noted that the Patient, Intervention, Comparison and Outcome (PICO) model is commonly used by medical students for the purpose of clinical questioning or searching. However, in the current study the PICO model is not considered because it is not the purpose of this study to evaluate diagnosis, treatment and intervention protocols. The motivation for this research study is to identify and analyse information search behaviour of medical students when searching across varying levels of task difficulty, and identify demographic variables that influence search behaviour when searching across varying levels of task difficulty.

LITERATURE REVIEW

Medical Students' Medical Search Behaviour

Mitchell et al. (1992) at a six month follow up study, found students who were trained to use *Grateful Med* performed medical searches more frequently in comparison to before being trained. Students were able to perform searches on overlapping topics and used Medical Subject Heading Terms (MeSH)[2] terms effectively. Students were also able to recover from search errors, zero hits and demonstrate the ability to broaden or narrow down a search. Shelstad et al. (1994) conducted another follow up study on the effectiveness of training medical students on *Grateful Med*. Unlike results of the earlier study (Mitchell et al. 1992), this follow up study was conducted when the same batch of medical student were in their 3rd year of study. Results of this follow-up study show that medical students were not able to use proper MeSH terms, used inappropriate sub-headings, misspelled terms and 20% of students received zero hits. Most students thought they were retrieving exactly what they needed but in fact they missed up to 100% of available citations and some students retrieved far too many results. The usage of *Grateful Med* also decreased to 57%, as medical students used it less than once a month. This indicates training only had a short-term benefit to improving a medical students' search session.

In a study comparing medical students and nurse practitioners search behaviour, Hersh et al. (2002) found that medical students spent more than 30 minutes on a task and they were only successful at correctly answering questions less than half the time. The ability to answer clinical questions with the aid of MEDLINE was low. The amount of time taken to conduct a search was typically longer than time spent on a patient. Hersh et al. (2002) concluded that students undergoing clinical training are at best moderately successful at answering clinical questions correctly. Clearly, medical students experience many search challenges when performing a medical search.

In another study, Auclair (2007) observed that medical students were able to correctly diagnose patients with the use of high order concepts. They were more successful at diagnosing a patient when they were able to identify the relationship between these concepts. Students who were not able to use high-order concepts or did not have the ability to relate these concepts were not able to properly diagnose a patient. Wildemuth et al. (1995) reported that there is little relationship in medical students' search behaviour between personal knowledge and search proficiency. Personal knowledge did not help students develop better search terms or increase search efficiency. In the same study, the authors investigated the type of questions asked by medical students with the use of clinical scenarios. Results show most questions asked by medical students only fell into five categories when the clinical scenarios presented required broader areas to be covered. First year medical students commonly asked identification type questions. Other students formed a question template and utilized it across clinical scenarios. These search traits indicate that medical students have limited ability to perform effective searches.

The Influence of Task Difficulty and Demographic Influence on Non-Medical Searches

In relation to non-medical type searching, task difficulty is found to influence the number of queries, the type of query issued, results viewing behaviour and timing aspects. Bystrom

[2]MeSH is a controlled Medical Subject Headings Index available at: http://www.nlm.nih.gov/mesh/MBrowser.html

(2002) found that more queries are issued when searching on a difficult task. Natural language type queries and query operators are often used when searching on a difficult task (Aula, Khan and Guan 2010). More sources and more content pages are viewed when searching on a difficult task (Li and Belkin 2008; Liu et al. 2012). In relation to timing, more time is taken to complete a difficult task (Li and Belkin 2008; Liu et al. 2010) albeit low navigation speed (Gwidzka and Spence 2006). Liu et al. (2012) also found most time was spent viewing results in comparison to issuing queries when searching in a difficult task. In relation to task difficulty and medical information searching, Inthiran, Alhashmi and Ahmed (2011) found medical students issue more and longer queries when searching on a difficult task. Apart from a research study conducted by these researchers, there is little information on how medical students search across varying levels of task difficulty.

Demographic variables have also been found to influence search behaviour. Hu, Zeng and Niu (2007) indicate gender and age influenced terms clicked on a page. Weber and Castillo (2010) found that education level influenced properties of the query and age influenced terms used in a query. Jones et al. (2007) found gender influence results clicking behaviour. While these research studies are based on non-medical searching, it is unclear if basic demographic variables (age, income bracket, gender) will influence medical searching behaviour.

Previous research studies provide us with information on the general search traits of medical students and highlights search challenges experienced by medical students. The research study conducted by Inthiran, Alhashmi and Ahmed (2011) provides us with basic understanding on the effects of task difficulty on querying behaviour. This research study is different in comparison to Inthiran, Alhashmi and Ahmed, (2011) because it focuses on the entire search session (querying and results viewing behaviour). Moreover an explanatory approach is used in an attempt to explain why medical students demonstrate particular search traits when across varying levels of task difficulty. We postulate medical specific demographic variables such as medical year of study, and medical search experience will influence search behaviour based on task difficulty.

OBJECTIVES AND METHOD

The objectives of this research are:

a) To study the effect of varying levels of task difficulty on search behaviour, and
b) To determine demographic variables influencing search behaviour when searching across varying levels of task difficulty.

The following research questions are developed for this study:

a) Does searching on varying levels of task difficulty influence the search behaviour of medical students?
b) Do demographic variables influence search behaviour of medical students when searching across varying level of task difficulty.

It is hypothesized that medical students do not demonstrate similar search behaviour based on task difficulty when searching on a medical task (as opposed to a non-medical task) and medical-type demographic characteristics influence search behaviour. The study focus specifically on task difficulty and demographic variables as it is believed these are amongst major influences on a medical students' medical search behaviour.

An interactive information retrieval experiment methodology is used in this research study. Unlike traditional experiment techniques, there were no treatment groups/non-treatment groups or dependent/independent variables. The purpose of this research study is to report on search behaviour and not on the effectiveness of a system or interface, as such there are no independent or dependent variables being declared. The experimental procedure consists of four data gathering techniques: a pre-experiment interview, search interactivity, structured observation and a post-experiment interview. The search interactivity data gathering technique involves the use of simulated situations and a personal task. Participants had to complete searching on three simulated situations (A,B,C) and on one personal task (P) (where applicable). Observation of search behaviour is segregated into three main sections: (a) querying behaviour, (b) search results interaction activity and (c) querying versus clicking behaviour and task completion time. These main sections represent typical segregation of a search session. A pilot test was conducted to test the interview questions, participants understanding of the simulated situations and the experiment methodology. A total of four (4) participants were part of the pilot test. As a result of the pilot test, simulated situations were refined and the interview questions fine-tuned. There were no new issues reported after the third pilot test participant. Thus we concluded the pilot test.

The main experiment involves a total of sixty (60) participants. Participants were recruited based on a single stage non-probability sampling method in a university setting. The stratification criteria of the study population are that participants had to be undergoing formal medical education. We recruited participants to the point of theoretical saturation in data analysis - where there was no more difference in observation of search behaviour amongst participants across simulated situations. The search domain chosen to study medical searching behaviour is *MedlinePlus*. All search sessions are restricted to *MedlinePlus* to avoid data interference or search behaviour influence as a result of searching on non-medical domains. Participants who have not used *MedlinePlus* were given some time to familiarize themselves with the web site. Participants are advised to search as normally as possible and were told stop the search once they have found relevant result/s or wished to terminate the search because they have not arrived at relevant result/s (unsuccessful search sessions - user-perceived satisfaction). Search history and cache is cleared from the web browser after each search task. The order of the personal task and simulated situations were rotated to reduce any effects of recency or primacy. Participants are handed the simulated situation one at a time and there is no time limit for each search task.

Participants who expressed interest to participate in this research study contacted the investigator and a meeting was set. The research experiment was conducted at the first researcher's office. Prior to the meeting, the first researcher informed participants to prepare a personal task for the experiment. At the meeting, participants were informed of the purpose of this research study. The researcher informed participants that the experiment was not to test their medical knowledge, expertise or about locating the correct answers to the scenarios. Participants were informed that their search activities would be observed and manually logged by the investigator. Participants were then given the Explanatory Statement and asked to sign the Consent Form if they agreed to participate in this research study. Next, the investigator administered the pre-experiment interview. After the pre-experiment interview the *MedlinePlus* web page was loaded using the *Internet Explorer Browser v7.0* or *Mozilla Firefox v3.0* depending on a participant's preference. The machine used to conduct the experiment is a standard computer with an Intel Core Duo 2 processor with a speed of 2.66 GHz and 1.96GB of memory. The machine is installed with *Windows XP Professional* operating system and other standard software

that are bundled with the installation. After completion of the first simulated situation, the post-experiment interview was conducted. Upon completing the post-experiment interview, the investigator provided participants with the next simulated situation. The same procedure was repeated for the subsequent simulated situation and the personal task. At the end of the experiment, participants were given a medium-sized notepad as a token of appreciation.

Pre-Experiment Interview

The pre-experiment interview is used to obtain demographic details, information on general search experience and medical search experience. Information on clarity of the simulated situations is also obtained at this stage. Participants' perception of clarity of simulated situations is obtained to determine that the descriptions provided in the simulated situations did not cause participants to experience any unnecessary 'search stress' due to poor construction of simulated situations. Participants rated their perception of simulated situation clarity using one of three options: clear, not clear or somewhat clear. When participants found the simulated situation to be unclear or somewhat clear, they were asked to provide some explanation as to why they found the simulated situations unclear or somewhat clear. Perception of task clarity was obtained before a search was conducted to ensure that the search experience did not influence participant's perception of task clarity. Participants who provided the researcher with their personal task were also asked to provide information on the search details of their personal tasks. Participants were inquired if they had searched for their personal task before and if so how much time was spent searching on their personal task.

Search Interactivity

To study interactive medical search behaviour amongst participants simulated situations and personal tasks are employed. Simulated situations are used because it provides a suitable substitute for real world information needs (Borlund 2000; 2003) to study interactive search behaviour. It also provides a platform to study and compare search behaviour amongst participants using a standardised scenario as it invokes a common information need amongst participants. Clinical and theoretical scenarios are developed to suite participants who would be at different stages of their medical study. Simulated situations are developed based on prevalent cases surveyed at a public hospital. These scenarios are then modified to fit the description of simulated situations. Each simulated situation has several search goals and covers different topics to provide search variety. Borlund (2000) used three simulated situations and one personal task in her research to test the effectiveness of simulated situations. Li et al. (2011) used three simulated situations for the purpose of measuring task difficulty. Poddar and Ruthven (2010) used three simulated situations and one personal task for the purpose of measuring emotional impact in a search process. Previous research studies have used three simulated situations for their respective aspects of study. Thus, based on previous research practise three simulated situations are developed for each category of users. Simulated situations A,B and C are provided in Appendix A.

The purpose of the personal task is to observe and compare differences in search behaviour when searching on the simulated situation and the personal task. The personal task is prepared by the participant. Each participant is supposed to prepare one personal task. The personal task could be a topic currently being searched, have yet to be searched for or searched before on a domain other than *MedlinePlus.* Out of sixty (60) participants only thirty eights (38) shared their personal for the purpose of this experiment.

Structured Observation

A structured observation technique is utilised for this research study because it allows for participants to explain their search behaviour and provide information on why they were dissatisfied with search results (user perceived satisfaction). In the querying activity section the following activities are observed: the number of queries issued per search session, the type of query issued (medical/non-medical), usage of query operators, length of the query (number of terms per query), query re-issues (exact), ineffective queries (query which returned no search results), queries without clicks (users did not click on search results) and usage of querying assistance. A query is qualified as a medical term query (medical query) when a term/s used in the query is found in the MeSH medical vocabulary database. The MeSH vocabulary database is used to qualify a query as a medical query because MeSH is the only formal controlled vocabulary thesaurus available managed by the National Library of Medicine.

In the search results evaluation activity section the following activities are observed: number of search results viewed for each task, result revisits, the number of sub-links clicked within the result page, viewing results beyond the 1st result page, using the Find Command (Control-F), and the numbers of unsuccessful search sessions (participants ended the search task without arriving at satisfactory results – user perceived satisfaction). In the third section, total querying versus clicking behaviour and task completion time is provided. Querying versus clicking behaviour provides the total number of queries issued (effective and ineffective) and results clicked (links and sub-links) for each search task. This is to provide a representation of querying and clicking activity in relation to task completion time. To measure theoretical saturation in data analysis, search behaviour demonstrated by participants is observed and recorded when searching across tasks. The open coding method (Strauss and Corbin 1990) is used for this purpose and no search actions were omitted. All activities performed by participants in search sessions are manually observed and logged. Analysis is performed after all sixty participants completed the experiment.

Post Experiment Interview

The post experiment interview took place after participants completed each search task. There are two parts to the post-experiment interview. The first part of the post-experiment interview is a structured interview relating to user-perceived subjective post-task difficulty (Li and Belkin 2008) of the simulated situations. Participants rated their perception of task difficulty using one of three options: easy, neutral (neither easy nor difficult) or difficult. Participants were also asked to rate their overall perception of task difficulty of the simulated situations against their personal task. This is to ensure that participants did not find the simulated situations too difficult or too easy in comparison to their personal task. This also provides the ability to compare search behaviour obtained from the simulated situation with search behaviour demonstrated when searching on the personal task. The second part of the interview is an unstructured interview. At this stage general feedback from the search session is obtained. Participants are encouraged to explain what search strategies worked/did not work, why they used certain search strategies and why they were dissatisfied with returned results (user-perceived perception – unsuccessful search sessions).

DEMOGRAPHIC DETAILS

Basic demographic details of participants are presented in Table 1. There were more female participants in comparison to male participants. The English Language is the second language for most participants.

Table 1: Basic Demographic Details of Participants.

Category	Details			
Gender (count)	Male		Female	
	16		44	
Use of English Language (%)	1st Language		2nd Language	
	40.0		60.0	
Nationality (%)	Malaysian	British	Middle Eastern	German
	65.0	15.0	15.0	5.0
Age (years)	Max	Min	Mean	SD
	26	19	22	2.1

Year of medical study and details of search experience are presented in Table 2. The number of participants is evenly distributed across the year of medical study. However, there were no participants from the fourth year of study. Participants search for medical information mostly on general search engines. Participants search for medical information to complete assignments or projects, to obtain more information after a lecture and to further understand information on diagnosis or treatment options for patients during clinical training. On average participants spend an hour a week searching for medical information. Demographic variables, search experience and medical year of study provided in Tables 1 and 2 are used to determine demographic influence on search behaviour. All participants have not undergone formal training in relation to medical information searching.

Table 2: General and Medical Search Experience of Participants and Year of Medical Study.

Category	Details				
General Search Experience (years)	Max	Min	Mean	SD	
	5.0	1.0	4.5	2.6	
Medical Search Experience (years)	6.0	1.0	3.5	1.2	
Medical Search Experience On General Search Engines Only (80.0% of total participants) [years]	2.0	1.0	1.7	0.9	
Medical Search Experience On Medical and General Search Engines (20.0% of total participants) [years]	3.5	1.0	3.6	1.2	
Years of Medical Study (%)	1st	2nd	3rd	5th	6th
	20.0	20.0	20.0	20.0	20.0

RESULTS

Five (5) participants expressed that the descriptions provided in the simulated situations was not clear. Amongst reasons provided for this is because participants "have not studied the topic" or "have not heard about a certain medical condition described in the simulated situations". There is no co-relation amongst issues experienced in relation to a specific

simulated situation. As such, unclarity, poor understanding or construction of the simulated situations did not affect participants' information search behaviour or perception of task difficulty.

Participants' perception of task difficulty is presented in Table 3. A Chi-Square test is performed to determine if there is significant difference in relation to task difficulty amongst the three simulated situations (A, B and C). A standardised residuals test is performed subsequently. Based on Table 3, a value of x^2=13.9, p=0.0076 is obtained for task difficulty. Results of the standardised residuals test indicate participants found simulated situation C to be more difficult (r=2.1) in comparison to simulated situation A and B. Results obtained from the standardised residuals test are used to analyse search behaviour based on difficulty level. It is acknowledged that perception of task difficulty is based on categorization of aggregated group perception rather than an individual's perception. This is done to capture the overall aggregated picture rather than a single participant's viewpoint which is bound to have variance and bias across participants (Hawthorne effect, interfering variable influence or lab type experiment setting).

Table 3: Participants Perception of Task Difficulty

Task\Difficulty	Easy	Neutral	Difficult
A	30	21	9
B	19	31	10
C	22	17	21

(Note: number denotes number of participants).

Search sessions are separated into three categories and presented in Tables 4, 5 and 6. Table 4 provides details on query details. Table 5 and 6 provide details on search results interaction activity and total querying versus clicking behaviour and task completion time respectively. Search session details (query length, number of queries, task completion time) are not normally distributed, thus the Kruskal-Wallis test is used to test for statistical significance. Mean rank values are provided along with the Kruskal-Wallis test results for the number of queries, query length and task completion time. Thereafter, an unplanned Dunn's post-hoc test is performed. For search results interaction activity, test statistic values are not significant at $p<0.05$ for results clicked and sub-links clicked. Therefore, mean or total values are used for discussion. Percentage values are reported for other aspects of query and search interaction activity because these interactions are not varied enough to be calculated based on test statistics. Percentage values are calculated based on the number of times these interactions are performed out of total search sessions. Results of the Kruskal-Wallis test indicate no significant difference amongst task for the number of queries issued. A value of H=9.33, p=0.0252 is obtained for query length. The longest queries are issued for simulated situation C and the shortest queries are issued for Task P. Results of Dunn's post hoc test indicate the median query length for all pair-wise comparison is different except for simulated situations A-B. Only the '+' query operator is used in all search sessions. Participants did not utilise any form of querying assistance in all search sessions.

Table 4: Querying Details of Participants.

Criteria\Task		A (N=60)	B (N=60)	C (N=60)	P (N=38)
# of queries issued per search session	Mean	2.1	2.3	2.1	2.0
	SD	1.5	1.7	2.2	1.1
	Min	1	1	10	1
	Max	8	9	1	6
	Median	1.5	1	1	2
	Test Statistics Mean # of Queries H=0.38 p=0.9443	113.3	106.7	107.9	110.4
Query length per search session	Mean	4.8	5.0	4.8	4.3
	SD	5.0	3.7	3.9	2.3
	Median	3	4	4	4
	Min	3	2	3	2
	Max	18	18	28	12
	Test Statistics Mean Query Length H=9.33 p=0.0252	111.2	111.7	122.4	83.0
% of ineffective queries	Spelling Error	8.0	2.0	4.0	0
% medical queries		0	0	0	31.0
% of query operator		0	2.0	0	5.0
% queries without clicks		28.0	4.0	23.0	7.0
% queries re-issued		0	0	3.0	0
% Querying Assistance		0	0	0	0

Table 5: Search Results Interaction Details for Participants.

Criteria\Task		A (N=60)	B (N=60)	C (N=60)	P (N=38)
# of search results clicked per search session	Mean	1.3	1.5	1.3	1.2
	SD	1.0	0.6	0.7	0.4
	Median	1	2	1	1
	Min	1	1	1	1
	Max	3	4	3	2
# of sub-links clicked per search session	Total	5	9	4	1
% of usage Control-F (find)		2.0	0	3.0	0
% of participants clicked beyond the 1st results page		3.0	10.0	12.0	7.0
% unsuccessful search sessions		5.0	10.0	15.0	11.0
Refining Search Results Assistance					
% of usage Refine by Keyword		0	0	3.0	0

Table 6: Total Querying Versus Clicking Behaviour and Task Completion Time of Participants.

Criteria\Task		A (N=60)	B (N=60)	C (N=60)	P (N=38)
Total Querying versus Clicking Behaviour		79 vs 75	87 vs 85	64 vs 70	51 vs 50
Task Completion Time Per Search Session (mins)	Mean	2.3	4.5	3.2	3.2
	SD	1.4	3.2	1.9	1.9
	Median	2	4	3	3
	Min	1	1	1	1
	Max	7	15	15	15
	Test Statistics Mean Task Completion Time H=17.92 p=0.0005	80.3	121.1	122.0	117.5

Twenty (20) participants perceived their personal task to be easier than the simulated situation. Seventeen (17) participants perceived their personal task to be of the same level of difficulty against the simulated situation and one participant perceived his/her personal task as harder than the simulated situation. These proportions provide an indication that participants did not find searching for the simulated situations too easy or too difficult. Twenty (20) participants have searched for their personal task before and the rest have not searched for their personal task. The longest a participant had spent searching on a personal task is two weeks and the shortest is a few days. Participants did not search for their personal task on a regular basis but do so when they had the time. The most number of sub-links was clicked when searching for simulated situation B. The most number of unsuccessful search sessions was experienced and the most results refining assistance was utilised when searching for simulated situation C.

A value of H=17.92, p=0.0005 is obtained for task completion time. The most amount of time is taken to complete simulated situation C and the least amount time is taken to complete simulated situation A. The median task completion time for all pair-wise comparison is different except between simulated situation B-C.

Search Behaviour Based on Task Difficulty

When searching on a difficult task, participants issued the longest queries and took the most amount of time to complete the task. They demonstrate high usage of results refining activities and experienced the most number of unsuccessful search sessions. They also viewed results beyond the 1st results page. Search behaviour demonstrated when searching on the personal task is similar to search behaviour demonstrated when searching on the simulated situations (with the exception of medical queries being issued on the personal task). Hence, search sessions demonstrated when searching for the simulated situations are a representation of participants 'real' search behaviour.

Participants utilise results refining assistance most often when searching on a difficult task. Participants took the longest time to complete searching on a difficult task but did not issue the most number of queries when searching on a difficult task. Instead, participants spent more time viewing results beyond the 1st page. Participants also experienced many unsuccessful search sessions when searching on a difficult task. Classification of search

behaviour demonstrated by participants when searching on a difficult task is provided in Table 7. The sub-classification of search behaviour demonstrated by participants when searching on a difficult task is known as active results locaters and long query issuers (viewing results beyond the 1st page, using the most results refining activities and issuing the longest queries), slow(most task completion time)and unproductive (experiencing the highest number of unsuccessful search sessions). Collectively, the search behaviour demonstrated by participants when searching on a difficult task is known as novice. This is because while dynamic querying activity is observed, the time taken to complete the task was slow and the search outcome was unproductive.

Table 7: Search Behaviour Demonstrated by Participants When Searching on a Difficult Task.

Search Behaviour Name	**Novice**	
Sub-Classification	Active Results Locaters and Long Query Issuers	Slow and Unproductive
Features	Longest queries Most viewing results beyond the 1st page Most results refining activities	Most task completion time Most unsuccessful search sessions

The post-experiment interview revealed that participants experienced many unsuccessful search sessions because they were unsure of the relevance of returned results. Participants did not click on returned results often but they did view results beyond the 1st page and used the most results refining assistance when searching on a difficult task. Participants reveal that they experience issues expressing queries and in determining suitability of returned results. However, they did not rely on querying assistance. No specific explanation is provided as to why participants did not utilise querying assistance. Participants viewed results beyond the 1st page, used results refining assistance often but did not click on result links. The post-experiment interview indicates participants evaluate returned results based on summary snippets only. The post-experiment interview also revealed that participants experienced issues in determining the suitability of returned results as they had to use results refining assistance often. Unfortunately, usage of results refining assistance was not helpful because participants felt that the results refining assistance was not providing them with streamlined results. While the querying versus clicking behaviour was even, the high task completion time indicates participants spent more time viewing results (viewing results beyond the 1st page but not issuing queries or clicking on result links). This indicates that there is a need to provide participants with better querying assistance. In addition, providing some indication on the relevance of returned results will encourage participants to click on more results. This will help to increase the number of results clicks in an effort to reduce the number of unsuccessful search sessions.

Demographic Influence on Task Difficulty

Twenty one (21) participants perceived searching for simulated situation C as difficult while the remaining thirty nine (39) perceived searching for simulated situation C as neither easy, neutral or difficult (neither). Based on this information, participants were manually separated into two clusters. Cluster 1 consisted of participants who found searching for simulated situation C difficult while participants in Cluster 2 found searching for simulated

situation C as neither (Table 3). Participants are first separated into clusters depending on the perception of task difficulty. Then, an ANOVA test is conducted to determine significant demographic variables that influence search behaviour. Thereafter a t-test is conducted to determine significant difference amongst clusters. For the t-test, the clusters were independent variables whilst the demographic characteristics were dependent variables

Based on Table 8, results of the ANOVA test indicates age is the only demographic variable that influenced participants in both clusters**.** However, results of the t-test indicate there was no significant difference in age for both the clusters. Thus demographic variables did not influence search behaviour when searching across varying level of task difficulty.

Table 8: Demographic Influence on Search Behaviour of Participants.

Demographic Details	ANOVA Test Results	T-test Results
Age	F=5.29, p=0.025	Not Significant

DISCUSSION

Results of this research study reveal that search behaviour demonstrated by participants when searching on a difficult task in most aspects resembles results of previous non-medical research studies (Brystom 2002; Li and Belkin 2008; Liu et al. 2010). Inthiran, Alhashmi and Ahmed's (2011) study reported that medical students issue the longest queries when searching on a difficult task. Results of the current study indicate only longer queries are issued with searching on a difficult task. While basic demographic variables influence general non-medical searching behaviour, neither basic nor medical type demographic variables influenced the search behaviour of participants when searching across varying levels of task difficulty. It is postulated that while demographic characteristics influence general search behaviour it is possible that demographic characteristics influence on search behaviour diminishes based on the property of the task (difficulty). This aspect certainly requires further investigation.

Results of this study indicate that there is no need to develop new search behaviour profiles based on description on search behaviour for participants based on task difficulty as search behaviour demonstrated based on task difficulty is similar regardless of the task type (medical/non-medical). The high usage of results refining activity, highest task completion time and highest number of unsuccessful search sessions indicate participants experience search challenges when searching on a difficult task. As such participants require specific information retrieval strategies to encourage high number of result clicks, reduce task completion time and reduce the number of unsuccessful search sessions. Thus the development of information retrieval strategies on medical domains have to take into account that medical students require specialised search assisting features.

CONCLUSION

This study utilised an interactive information retrieval methodology to analyse the search behaviour of medical students when searching across varying levels of task difficulty, and to ascertain if demographic variables influence search behaviour when searching across varying levels of task difficulty. The theoretical contribution includes better understanding

of search behaviour and the practical contribution better inform the need to design user centric information retrieval strategies. Existing data can be used to determine what type of queries is issued by typical medical searchers when performing a medical search. There is an opportunity to examine search patterns of successful and unsuccessful search sessions. This will allow for the identification of search behaviour patterns when searching on a successful and unsuccessful search session.

This study provides many interesting findings; however there are also limitations that should be acknowledged. A total of thirty eight (38) personal tasks were provided by medical students. However, the number of personal tasks provided by medical students were more than thirty (more than half the number of participants), thus provides sufficient data for analysis. Furthermore, not all levels of task difficulty were available for the categories of typical medical searchers. Results obtained from this research study cannot be generalised. It is limited to the study population and the domain used to perform this research study. In this research study, only task completion time is analysed. Time spent on each result page clicked and time spent viewing returned results (without clicking) was not taken into account. Time spent on individual results page and time spent viewing results (without clicking) will allow for better understanding of how task difficulty influence results viewing activity. We acknowledge that there is a possibility that interfering influence could also affect search behaviour however we limit this condition by accounting aggregated group perception of task difficulty instead of individual perception of task difficulty.

ACKNOWLEDGEMENT

The authors would like to thank medical students from the University of Malaya Medical Centre (UMMC), Monash University Malaysia, Malaysian Allied Health Sciences Academy (MAHSA) and medical students from England, Ireland and Egypt who were undergoing elective clinical training at UMMC.

REFERENCES

Abderrahim, Q, Jean-Paul, F., Jean-Louis G., and Marc, B. 2015. Information-seeking behavior during residency is associated with quality of theoretical learning, academic career achievements and evidence based medical practice: A Strobe-compliant article. *Medicine*, Vol. 94, no. 6: 535-545.

Auclair, F. 2007. Problem formulation by medical students: An observational study. *British Medical Education,* Vol 17, no. 6. Available at doi:10.1186/1472-6920-7-16.

Aula, A., Khan R.M., Guan, Z. 2010. How does search behaviour change as search becomes more difficult? *Proceedings of the 28th Human Factors in Computing Conference (CHI 2010*), pp.35-44. New York, NY, USA: ACM

Borlund, P. 2000. *Evaluation of interactive information retrieval systems.* Unpublished Thesis, Abo Academy University Press.

Borlund P. 2003. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems, *Information Research*, Vol. 8, no. 38, paper no 152 Available at: http://information.net/ir/8-3/paper152.html.

Bystrom, K. 2002. Information and information and sources in tasks of varying complexity, *Journal of the American Society for Information Science and Technology*, Vol. 53, no.7: 581-591.

Can, A.B. and Baykal, N. 2007. MedicoPort: A medical search engine for all. *Computer Methods and Programs in Biomedicine*, Vol. 86, no. 1: 73-86

Gormley, G., Bickle, I., Thomson, C., Collins, K. 2009. Online learning in clinical skills: the Belfast experience, *The Clinical Teacher*, Vol. 6, no. 1: 46–50

Gruppen, L.D., Rana, G.K. and Arndt, T.S. 2005. A controlled comparison study of the efficacy of training medical students in evidence based medicine literature searching skills, *Academic Medicine*, Vol. 80, no. 10: 940-944.

Gwizdka, J., and Spencer, I. 2006. What can searching behaviour tell us about the difficulty of information tasks? *Digital Library of Information Science and Technology.* Available at http://dlist.sir.arizona.edu/1818/01/Gwizdka%5Fwhat.pdf.

Hersh, W. 2009. *Information retrieval: A health and biomedical perspective.* 3rd. ed. New York, NY, USA: Spinger- Verlag.

Hersh, W.R., Crabtree, K.M., Hickam, D.H., Sacherek,Y., Friedman, C.P., Tidmarch, P., Mosbaek C., and Kramier, D. 2002. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association*, Vol. 9, no. 3: 283-293.

Hu, J., Zeng, H.J., Li, H., Niu, C. and Chen, Z. 2007. Demographic prediction based on user's browsing behaviour, *Proceedings of the 16th International World Wide Web Conference (WWW 2007),* pp.151-160. New York, NY, USA: ACM.

Inthiran, A., Alhashi, S.M., and Ahmed, P. K. 2011.Investigating interactive search behavior of medical students: An exploratory survey. *Australasian Conference in Computer Human Interaction,* Canberra, Australia, 28 Nov – 2nd Dec 2011.

Jones, R., Kumar, R., Pang, B., and Tomkins, A. 2007. I know what you did last summer: query logs and user privacy. *Proceedings of the 16th International Conference on Information and Knowledge Management CIKM (2007)*, pp. 909-914. New York, NY, USA: ACM.

Li, Y. and Belkin, N.J. 2008. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, Vol. 44, no. 6: 1822-1837

Li, Y., Yu, C., Liu, J., Cheng, Y., Wang, X., Chen, P. and Wang, Q. 2010. Measuring task complexity in information search from user's perspective, *Journal of The American Society of Information Systems and Technology*, Vol. 48, no.1:1-8

Liu, C., Liu J,.Cole, M., Belkin, J. and Zhang, X. 2012. Task difficulty and domain knowledge effects on information search. *Proceedings of the American Society for Information Science and Technology ASIS&T Conference*, pp.142-149

Llic, D., Tepper, K. and Misso, M. 2012. Teaching evidence based medicine literature searching skills to medical students during the clinical years: a randomized controlled trial. *Journal of Medical Library Association*, Vol. 110, no. 3: 190- 196

Mitchell, J.A., Johnson, E.D, Hewitt, J.E, and Proud, V.K. 1992. Medical students using grateful Med: Analysis of failed searches and a six month follow-up study. *Computers and Biomedical Research*, Vol. 25, no. 1: 43-55.

Mitchell, J.A., Johnson, E.D. and Proud, V.K. 1990. New thought about medical students as effective searchers of MEDLINE. *Academic Medical*, Vol. 65, no. 7: 434.

Nadkarni, P. M. 2000. Information retrieval in medicine: Overview and applications, *Journal of Postgraduate Medicine*, Vol. 46: 116-122.

Poddar, A and Ruthven I. 2010. The emotional impact of search tasks. *Proceedings of the 3rd Information Interaction in Context (IIiX) Conference 2010*, New Brunswick, NJ, USA, August 18 - 21, 2010, pp.35-44. New York, NY, USA: ACM.

Peterson, M.W.,Rowat, J., Kreiter, C., Mandel, J. 2004. Medical students' use of information resources: Is the digital age dawning? *Journal of Academic Medicine,* Vol. 79, no. 1: 89-95.

Schilling, K., Wiecha, J., Polineni, D. and Khalil, S. 2006. An interactive web-based curriculum on evidence based medicine: Design and effectiveness. *Journal of Family Medicine,* Vol.38, no. 2: 126-32.

Schurtz, S. 2009. Thinking outside the classroom: Providing student-centered informatics instruction to first- and second-year medical students, *Journal of Medical Reference Services Quarterly*, Vol. 28, no. 3: 275-281

Shelstad, K.R. and Clevenger, F.W. 1994. Online searching strategies of third year medical students: Perception. *Journal of Surgical Research,* No.56, no. 4: 338-344

Strauss, A and Corbin, J. 1990*. Basics of qualitative research: Grounded theory procedures and techniques.* CA: Sage Publications.

Webber, I. and Castello, C. 2010. The demographics of web search. *Proceedings of the 33rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, 19-23 July 2010, pp.78-82.

Wildemuth, B.M., Bliek, R.D and Miya, T.S. 1995. Information seeking behaviors of medical students: A classification of questions asked of librarians and physicians. *Bulletin of the Medical Library Association,* Vol. 82, no. 3: 295-304.

APPENDIX A

Simulated Situation A:

Simulated work situation: As part of your study of medicine it has been discussed why arterial blood gases (ABG) test are important. Now you want to learn more about this ABG, why is it important and how is this test conducted. Indicative Request: Find for an instance, information about why do we need to perform ABG, how would you conduct an ABG test and what is included in an ABG.

Simulated Situation B:

Simulated work task situation: As part of your medical training, you have been tagging a senior doctor in the accident and emergency ward in a hospital. Yesterday, you observed a case where a mother brought in a toddler who appears to be choking but conscious. You were curious of the situation and would like to use MedlinePlus to learn how to treat the toddler according to the symptoms described. Indicative Request: Find for an instance, information to treat the toddler according to the symptoms mentioned above and to learn more about the toddler's condition.

Simulated Situation C:

Simulated work task situation: As part of your medical training, you have been asked to complete morning rounds in a hospital ward. A medical officer (MO) doctor presents to you a case where a patient is recovering from acetaminophen or paracetamol poisoning. The MO has asked you to find out possible ways to best treat this patient and to report your findings the following day. Indicative Request: Find for an instance, information on MedlinePlus to help you come up with step/s to treat the patient.