Exploring Information Interactions
in the Context of Google

Frances Johnson
Department of Languages, Information & Communications, Manchester Metropolitan University, Geoffrey Manton Building, Manchester M15 6LL, UK. E- mail:
F.Johnson@mmu.ac.uk

Jennifer Rowley
Department of Languages, Information & Communications, Manchester Metropolitan University, Geoffrey Manton Building, Manchester M15 6LL, UK. E- mail:,
J.Rowley@mmu.ac.uk

Laura Sbaffi
Department of Languages, Information & Communications, Manchester Metropolitan University, Geoffrey Manton Building, Manchester M15 6LL, UK. E- mail:,
L.Sbaffi@mmu.ac.uk

The study sets out to explore the factors that influence the evaluation of information and the judgments made in the process of finding useful information in web search contexts. Based on a diary study of two assigned tasks to search on Google and Google Scholar, factor analysis identified the core constructs of content, relevance, scope and style, as well as, informational and system 'ease of use' as influencing the judgment that useful information had been found. Differences were found in the participants' evaluation of information across the search tasks on Google and on Google Scholar when identified by the factors related to both content and ease of use. The findings from this study suggest how searchers might critically evaluate information, and identifies a relation between the user's involvement in the information interaction and the influences of the perceived system ease of use and information design.

Introduction

Information seeking is ubiquitous, an everyday activity, with a generation raised to 'Google' it, find it and expect instantaneous information. Searching for information is variously described as 'easy' to 'difficult' and often frustrating. One reasoned explanation is that the extent of the challenge of finding information depends on the search task. Searching for something on a general subject or a popular query, such as the birthdate of a famous person, is very different from an open ended research query or exploration of a previously unknown topic. The sea of information makes it easy to find something or anything on a given topic but, on the other hand, when we do not regard 'satisficing' as appropriate, it can be very difficult to find *relevant* information. The concept of relevance is central to the Information Sciences (Saracevic, 1975) and in the context of searching for information it can be somewhat generally, but usefully, conceptualized as the perception that the content of the information retrieved directly bears upon their

Corresponding author. Frances Johnson

given information need or interest. In short, finding relevant information requires the searcher's involvement in making some critical judgment. The fundamental importance of relevance to successful information searching, beyond finding sufficient information, has triggered a resurgence of interest in the concept of information literacy as an important problem solving skill in identifying relevant information to resolve some identified information need. The Association of College & Research Libraries (ACRL) Information Literacy Competency Standards for Higher Education directly quotes the American Library Association in defining information literacy as a set of abilities requiring individuals "to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (ALA, 1989). The core competencies focus on the critical evaluation of information from a variety of sources, based on a recognised information need. Empirical studies of information search (e.g., Blandford & Attfield, 2010; Walraven, Brand-Gruwel & Boshuizeet, 2009) have stressed the pivotal importance of evaluation in finding useful information and define the process as one in which information is judged to be reliable, credible and topical (relevant) and, through integrating this information with our prior knowledge, we learn, in the process of its acquisition, for the task in hand. Yet evidence in the literature focusing on students suggests that they may lack the necessary critical processing and evaluation skills when searching for information (Bilal, 2000; Brown, 2001; Coiro, 2003; Leu, 2000; Pritchard & Cartwright, 2004). Fosmire and Macklin (2002) write that students believe they already have information skills because they can find something on Google to answer almost any question, but they are unable to critically evaluate that information at a deeper level. They suggest that what has been acquired is a confidence in the ability to search but also needed is the capacity for critical reflection on and, evaluation of, the information found.

The present study sets out to explore the judgments made in the critical evaluation of information, specifically the factors that may influence judgments in finding useful information. With respect to describing information seeking behaviour we focus on the goal driven information seeker who, in context of a recognised need, deconstructs the quest into a series of subtasks including the specification of a set of search terms and, in interacting with the information retrieved, performs some evaluation of the information to terminate the search (or, in an interactive search, as part of an interim feedback process). The focus is, however, on identifying the constructs of the information evaluation as an activity in the search context, placing the judgments of relevant and useful information as the core cognitive activity in the information interaction. The assumption is that, in the effective and dynamic search, the information retrieved is assimilated by the searcher to form a personal perspective on the information retrieved and to judge its relevancy. To explore how people evaluate information in search contexts, the approach taken is to identify the factors that may influence these critical judgments. Furthermore, to obtain deeper insights, the study investigates the effect that the search task might have on the level of the searcher's involvement in the evaluation of the information. This study of the factors influencing cognitive activity, specifically in search contexts, differs in one important respect from broader studies of information seeking processes where the competencies in the critical evaluation of information may be developed over time (e.g., as in Kuhlthau, Heinstrom, & Todd's (2008) study of the stages of students' information seeking) and which is generally regarded as unchanging regardless of the technology. In the context of a web search, the technology, or the user's perception of the effectiveness of the system used to carry out the search, is considered alongside the information itself as influencing the evaluation of the usefulness of the information retrieved. Given that it is now well accepted that university students are drawn to Google for its convenience and self- explanatory interface (Mullen & Hartman, 2006), this perceived ease of use may impact on the assessment of the information retrieved. Indeed the basis for much of the criticism of Web search engines for making search 'too easy' is that too often we are prepared to substitute convenience and a reliance on the effectiveness of the engine for the more demanding critical evaluation of the usefulness of the information found.

Study Aims

This study aims to explore the critical judgments formed when searching for and evaluating information in response to a task based query. Based on previous research, the critical judgments formed in evaluating the information are not fixed or static (for example, see Kelton, Fleischman and Wallace's (2008) definition of 'trust' formed on digital information, or Borland (2003) on the dynamic nature of relevance judgments) but may be identified by their influencing factors. It is also assumed that the extent of the searcher's involvement in the interaction is partly determined by the requirements of the search task (for example, Tom, Freund, Kopak and Bartlett's (2002) model of information interaction identifies a process affected by the user, the system and the content) and as such, the judgments formed in the evaluation of the information, may vary across search tasks. Previous research identifies the possible factors that influence these judgments and, to help further investigate this cognitive activity in the search context, these are identified in this study as either information attributes (such as, credible and/or relevant) or contextual factors, such as a perceived ease in

finding the information. In searching for information, and in the ensuing evaluation of the usefulness of the information found, it is likely that we are influenced by the perceived effectiveness of the system or engine used and, possibly, by the support we feel we have experienced in finding the information, that is the system ease of use. Specifically, in information seeking, 'ease of use' may be attributed to the system (its effectiveness as well as the tools it provides to carry out the task) as well as to the interaction with the information during the search task. This study is based on the following three premises that provide the framework for the investigation into the critical evaluation of information in search:

(1) In looking for information students will evaluate the information ('usefulness') on the criteria of credibility/reliability/quality, relevance, style and scope.

(2) In making this judgment students will be further influenced by the perceived ease of use of the system in retrieving information as well as by the support experienced during the process of searching for the information.

(3) The search task will affect the students' involvement in the evaluation of the information as evidenced by the types of influencing factors and their relationship to the judgments formed.

The empirical study was conducted to confirm the most important factors in determining the information evaluation and to identify any impact of the search task on the interaction with the information. The diary study employed was pseudo or artificial as the participants were assigned tasks to carry out in their own time and report back on their perception of the information found and on its evaluation. Specifically they were asked to respond to the possible factors influencing their judgment of usefulness, which were both 'on the page', relating to the content of the information, and 'off the page' relating to the perceived ease in finding the information. Thus the pseudo diary study seeks to identify the criteria on which students (as the participants in the study) judge the information retrieved and whether differences can be discerned in the judgments made across tasks involving the user in the search. The premise that the critical judgments formed in evaluating information are determined by the influencing factors allows differences in these judgments to be discerned. Understanding the cognitive activity of evaluating information in search contexts places a particular interest in the relation between the judgments formed and the contextual factors, such as the perceived effectiveness of the system.

Theoretical Context

This study poses the question – 'What does it mean to be information literate in the specific context of web search?' and sets out to investigate the cognitive activity associated with evaluating information to successfully complete a task. Insight into this activity is obtained through previous research that identify the constructs that contribute to the evaluation of information, such as relevancy, credibility, reliability, style and scope as well as the research which suggests that, at times, evaluation takes into account perceptions of ease of use and the experience of the process in finding the information.

In traditional information retrieval theory, central to the critical evaluation of information is the concept of relevance which refers to the extent to which a piece of information directly satisfies an information need. Subjective relevance refers to the 'aboutness' of the document as perceived by the user in relation to the information need: if the user believes a document is about the topic of interest, then for them it is topically relevant. Going beyond subjective relevance, cognitive relevance refers to the relation between the state of knowledge/ information need and texts retrieved, whereas situational relevance refers to the usefulness or appropriateness of information in the resolution of a problem. In this continuum, relevance is thus conceptualised as a dynamic judgment (Saracevic, 1975; Borlund, 2003) in which the subjective judgment of the information 'topicality' may be regarded as a basic requirement of the information as perceived by the user in relation to an information need, and its 'usefulness' with regards to its actual use in directly resolving the given problem. In this study we focus on the judgment of usefulness which may be preceded by a judgment of topical relevance and influenced by a range of independent variables. Many empirical studies have been carried out to identify and understand the criteria for relevance judgment, typically reporting influencing factors, including style, level of difficulty, recentness, accuracy, reliability (e.g., Schamber & Bateman, 1996; Wang & Soergel, 1998; Hirsch, 1999). Further and more recent studies, confirming these factors, have investigated how people search, and identify the core judgments of the information as reliable, credible and relevant often across domains. Toms, Freund, Kopak, and Bartlett's (2003) studies in consumer health, shopping, travel, and general research domains suggest strongly that in information seeking we evaluate information on these core criteria. Tang, Shaw Jr.and Vevea (1999)

suggest that the content covered, that is the breadth and depth (or scope) of the information are appropriate with regards to the user's need. Xu and Chen (2006) identified the criteria of topicality and novelty as essential alongside understandability and reliability. Novelty refers to the extent to which the information is informative or new to the user thus producing cognitive change helping to characterise relevance as a subjective and also a dynamic concept. Balatsoukas and Ruthven (2012) observed participants' eye movements and relevance judgments in the context of web search and confirmed topicality as the core criterion but also identified scope, tangibility and a new dimension of serendipity, or serendipitious judgments, being purposefully used when participants had not found relevant information in the first few results in a ranked list. Further, core judgments in the evaluation of the information may be based on the user's verification or checking of the content with other sources. Rieh and Hilligoss (2008) confirmed that students undertook evaluation of information based on quality control mechanism (refereeing, editing), and verification (through using multiple sources and co-referencing) as well as on their current knowledge on the topic. Hargittai, Fullerton, Menchen-Trevino and Thomas (2010) on the basis of a study of first year undergraduate students, suggested that the most important influences in credibility assessment were sources for validation such as whether fact or opinions are presented, authorship, and linking sites. Walraven, Brand-Gruwel and Boshuizeet's (2009) study examines how secondary school students evaluate information and sources when searching the web, and identifies three clusters of factors that they use in evaluation as reliability, verifiability and usability. Another factor that emerges is information presentation. Sillence, Briggs, Harris and Fishwick's (2007) diary study of how patients evaluate and make use of online health information, found that the factors contributing to the selection and trust of web sites can be divided into content factors (e.g., informative content, unbiased information, clear, simple language) and design factors (e.g., clear layout, good navigation aids, interactive features).

Past research into relevance judgments has identified and confirmed core criteria to be topicality, credibility, reliability and tangibility, tangible although the actual use of these judgments is rarely studied. Furthermore consideration is not given to the dynamic nature of the relevance judgment when made in interactive search contexts ranging from a superficial check that the information looks good, to a judgment of topicality, to involving the user in the evaluation and assimilation of the information in cognitive relevance and, ultimately, the interpretation and application of the information in situational relevance. In making these types of relevance judgments the searcher is increasingly involved in creating a perspective on the information and, it may be assumed, changing their cognitive state as they become informed in the process. This raises the question, does the criteria for evaluating the information for relevance vary depending on where on this continuum the user may be placed? This is difficult to discern in an empirical study. However, research has suggested that people may select documents as relevant based on some blind trust in the search engine ranking algorithm (Pan et al., 2007). Is it possible that this reliance only occurs when the participant is not actually required to be involved in cognitive or situational relevance to complete the task? Google is consistently reported to be the search engine of choice and it would appear that users believe that certain search engines will find 'the best' results. Jansen (2008) found that users preferred one set of results over others, and claimed them to be superior, even though they were shown the same exact set of results in both content and presentation. Perhaps not surprisingly, Helms Park, Radia, and Stapleton (2006) found that results yielded by Google Scholar (GS) were rated superior to those from Google in the categories of authority and reputation, objectivity, academic rigor and transparency, although they did not find these differences in the rating between GS and the library e-resources.While core criteria can be identified for judging the document, to gain insight specifically into the relevance judgments made it is important that the influencing factors are further identified. These, as Sillence, Briggs, Harris and Fishwick (2007) suggested, may relate to content or design. In web search these design factors may be recognised as a perceived ease of use of the system and may include interactive features which enable the user to become more involved in evaluation the information, for example in actively verifying the information, or the design factors which affect the ease in which the user forms relevance judgments. Thus the framework for this investigation should identify searchers' involvement in judging the information as well as the influencing factors in the searchers' evaluation of the retrieved information and distinguish these as content factors and those that are peripheral to the content judged. The approach taken was to develop a psychometric questionnaire to measure the constructs such as relevance and credibility, as well as perceptions of other factors, such as system effectiveness, in the evaluation of information.

Methods

This study is based on pseudo diary study to obtain participants' assessment of the retrieved results from two assigned search tasks, one on Google and the other on Google Scholar. A questionnaire was used to obtain a judgment of the usefulness of the information found and to assess the factors that may have contributed to this judgment. The participants were students and the tasks were tailored to find information to help prepare an assignment with the query types suited to the particular engine. The query put to Google was of a 'general type', to find information on Alan

Turing and his contribution to the development of Computing, whilst the 'informational' query put to Google Scholar was more open ended to find background information on the topic of young people's use of Short Message Services (SMS) and its effect on their written language. These were chosen to emulate a likely search context in which students might recognise a need for information to complete an assignment. A Google search is likely to be the first port of call and, whilst millions of results will be retrieved for a broad query such as 'Alan Turing Computing', the student is likely to notice that the top results include links to online encyclopaedias and to trustworthy sites such as History from the BBC. Thus, with a simple search on Google, the student will find information linking Turing to the theory of Computation and Artificial Intelligence and may pick up further names, landmarks and concepts to look out for in further information gathering for the essay. In this task, we expect our participants to find the information they might judge to be useful from an initial query on Google. The informational query, on young people's use of SMS, on the other hand, is best resolved by some interaction taking place whereby the searcher conducts the search iteratively learning from the results to tailor the query towards the specific topic. Again we might expect our participants to initiate the query using the terms in the task 'sms young people written language' or 'sms effect written language' to scope what is available. They might find a useful paper on the first go, but it is important to note the scholarly nature of the articles returned on Google Scholar which may prompt an initial exploration to lead to further searches using acquired terms (for example, sociolinguistics, orthography). Thus, we might expect the student to evaluate the gathered information and formulate a more personalised perspective to investigate the subject. For this reason, and in the context of the diary study, we assume that the participant will have worked a little harder at formulating this query to obtain a few useful results to satisfy the task set.

Implementation

Sampanes, Snyder, Rampold-Hnilo and White (2011) identify the use of the diary for data collection as it allows participants to explain their behaviour and experiences and for the researcher to gain insight on the impact of key situations or events. Unlike fixed questionnaires or laboratory settings, a diary can be taken away and completed whilst carrying out some task or when using the technology. As our participants were required to work on a given task rather than report on an event naturally occurring in the specified time we refer to this study as a pseudo-diary study. There is, of course, a risk on the reliance on self-reporting in the diary study, however, the distribution of the task in class was deliberate in an attempt to motivate participation in the project. Participants were undergraduate and postgraduate students on courses in Information and Communications in the UK. Completed questionnaires were collected by the researchers at the start of the same class the following week.

Data Collection

The questionnaire is presented in Appendix 1. Prior to starting the search, respondents were asked about their certainty in finding the information and about their familiarity with and knowledge of the task topic. After conducting each search, participants were asked to judge, on a seven point Likert scale, the usefulness of the information found with regards to the task (USE) and their certainty that appropriate information had been found (CERT). In order to collect responses reflecting a deeper evaluation of the information, participants were asked to think about making use of the information and respond to items relating to their certainty that appropriate information had been found, for example *I would use this information to work on the task in hand*, and *If asked I think that other people would value this information*. To obtain a measure of cognitive relevance, as distinct from usefulness, participants were also required to think about making use of the information and to respond to items such as, *the information was just the right amount for me*, *I expect to feel well informed on this topic*, *The information has helped me to learn about the subject*, and *The information would help complete the assignment* in an attempt to capture the users' forming a personal response to the information.

They were then asked to respond to Likert-style statements (the items) designed to investigate their judgment of the information found and based on the aspects and possible latent constructs of content, relevance, style, scope, information and system ease of use. These constructs (as shown with their items in Table 2) were labelled Content (CON1-4), Relevance (REL1-3), Style (STY1-3), Scope (SC1-3), Informational Ease (INF1-5) and System Ease of Use (SYS1-4) and, informed by previous research, the statements were largely drawn from a rephrasing of possible items to reflect the construct and as used in the literature. For example, to measure relevance assessment we asked survey participants their level of agreement with the statements that *the content is related to the topic of interest* (this being item REL1) and *that the information tells them most of what they need to know* (item REL3). In this way we avoid asking the participants to directly score the information on the construct of 'relevance'. Instead their judgment of

relevance may manifest through responses to the statements. Responses were requested to items relating to aspects of the information content, style, scope such as *The information is accurate* (item CON2)*, I can believe the content (*CON1*), The information appears authoritative (*CON3*), The information is well written* (STY3*), The content is easy for me to understand (*SC3*)*. To measure a construct we might label as 'informational ease' respondents were asked to think about finding the information and to respond to items such as, *It was easy to identify what the information was about* (INF1) and *I felt that my query had been understood* (INF2). A measure of the system perceived ease / effectiveness was based on responses to items such as, *The system is good at finding information that matches my query (*SYS2). All codes for each item used in the questionnaire are shown in Table 2. Content validity was grounded in the use of previous research to inform the items used in the survey. For face validity the researchers discussed the phrasing of the questions and ran a pilot on which minor changes were made in the phrasing of the questions. Finally, in the questionnaire, a further question was included to gather qualitative insights and requested the participants to write down the three factors which most influenced their judgment of the usefulness of the information for each task.

Data Analysis

Data were entered into IBM SPSS Statistics. A total of 102 students participated: male (50), female (52) and the majority in age range 18-22 (82). The questionnaires were distributed in selected classes running during induction week reaching, about, one third (30-40%) of the population (students studying on the taught programmes in the department) and were collected the following week with a near 100% response rate. By level, we obtained a fairly even distribution of undergraduate first year (29), second year (24), third year (33) and postgraduate (16). The background profile of the student's certainty in finding the information and on their familiarity and knowledge of the task topic is presented in Table 1. Based on the mean scores obtained from these questions prior to conducting the search, Postgraduates and Year 1 reported the highest level of certainty that the information can be found, with Years 2 and 3 reporting a slightly lower scoring in response to the question. The level mean for certainty dropped slightly across the years from the Google to Google Scholar (GS) tasks. This drop was also observed for the question *How certain are you that you will find the information*? and, in the responses to the statement relating to search experience, *I am used to searching for this type of query* with lower mean scores obtained across the groups when asked in the GS task context. The trend was reversed for the questions relating to familiarity with and knowledge of the topic with the student level mean higher for the GS task than with the Google task on Alan Turing. Overall participants reported a high level of certainty that the information can be found (5.54 Google, 5.49 GS) and a rather cautious mid rating of their knowledge on the topic (3.3 Google, 3.78 GS). When asked about the search and in general how satisfied they were with the information found after the initial query, a mean of 5.31 was obtained for the Google task and a mean of 4.74 for the GS task. However, only a small number reported having revised or refined their initial query when asked and, in the absence of this further data, we can assume that the majority found the information they judged to be useful on the results from the initial query.

Table 1: Profile of students' experience with and confidence in the search tasks.

		UG yr1	UGyr2	UGyr3	PG	Total
How certain are you that the information can be found	**Google**	5.66	5.42	5.30	6.00	5.54
	GScholar	5.79	5.29	5.24	5.75	5.49
How certain are you that you will find the information?	**Google**	5.79	5.42	5.27	5.63	5.51
	GScholar	5.45	5.29	4.94	5.19	5.21
I am used to searching for this type of query	**Google**	4.90	4.63	4.79	4.50	4.74
	GScholar	4.45	3.75	4.21	4.19	4.17
How familiar are you with the task topic?	**Google**	3.83	3.33	3.67	2.81	3.50
	GScholar	4.90	3.87	4.09	3.13	4.12
How knowledgeable are you about this topic?	**Google**	3.55	3.21	3.64	2.31	3.30
	GScholar	4.28	3.54	3.82	3.19	3.78
If asked, how willing would you be to inform others about this topic	**Google**	3.83	3.54	3.73	3.00	3.60
	GScholar	4.03	3.83	3.85	3.44	4.78

Table 2 presents the descriptive statistics, the means and standard deviations, for all the items relating to the assessment of the information usefulness on the intended constructs of Content, Relevance, Style and Scope, Information and System Ease of Use. When asked to think about the information judged to be useful to the task, the

highest mean relating to Content was on the item *the content is accurate* CON2(5.76) on Google and CON2(5.49) on GS. The next highest was *believe content* CON1(5.52) on Google and, on GS, good *quality information* CON4 (5.48). On Relevance, the highest means were on *the content is related to topic*, again with Google obtaining a slightly higher mean, REL1(5.56) compared to GS REL1(5.27). On Style, the highest mean on Google was obtained in rating the *information is structured* STY1(5.51)) and on GS the highest rating was on the item, *is well written* STY2(5.63). The assessment of information Scope was highest in responding to the *information is easy to understand* (SC3). With regards to the view held on System-Ease of Use, the highest mean obtained for Google was on the item *I can work the system* SYS1(5.8), also high means were obtained for the *system is good* SYS2(5.56) and *works well* SYS3(5.55). Google Scholar obtains slightly lower means across these items at SYS1(5.45), SYS2(5.32) and SYS3(5.24) and, although possibly indicating a lower certainty in the assessment of the engine, these are still the highest scoring items when it came to rating Google Scholar. On the construct of Information Ease the highest agreement was reported on the item *ease in identifying what the retrieved information was about* INF1(5.68) on Google and INF1(5.24) on GS. With probability at <.005 paired t tests indicate that the participants' rating of the information usefulness (USE) increased significantly on Google (t(101) = -2.907, p= .004). Differences in the ratings on the above items (relating to Content, Relevance, Style, System Ease of Use) were not significant when compared across the Google and GS results. Although, at p<.05, tests did show a significant difference in the increased rating obtained on the Information Ease item '*it was easy to identify what the info was about'* (INF1) (t(101) - -2.122, p=.036).

Table 2: Means levels for intended constructs with paired samples test on items rated highest.

		Turing search (**Google**)		SMS search (**Google Scholar**)		GS and Google t-tests
Thinking about the information you chose as being the most useful for the task						
		Mean	**SD**	**Mean**	**SD**	
I can believe the content (CON1)		**5.52**	1.79	5.41	1.46	
I think the content is accurate (CON2)		**5.76**	1.28	**5.49**	1.06	-1.959, p=.053
The information appears to be authoritative (CON3)		5.38	1.72	5.43	1.55	
The information is of good quality (CON4)		5.43	1.59	**5.48**	1.28	
The content is totally related to the topic of interest (REL1)		**5.56**	1.24	5.27	1.31	-1.767, p=.080
This subject area is only marginally related to the topic (REL2)		4.04	1.93	3.73	1.79	
The information tells me most of what I need to know (REL3)		5.42	1.23	5.20	1.21	
The information is too detailed (SC1)		4.07	1.90	4.14	1.61	
The coverage is too general for me (SC2)		3.91	1.62	3.75	1.57	
The content is very easy for me to understand (SC3)		**5.35**	1.33	**5.20**	1.25	
The author/org appears to be knowledgeable (STY1)		5.37	1.65	5.41	1.56	
The information is clearly structured (STY2)		**5.51**	1.43	5.40	1.45	
The information is well written (STY3)		5.38	1.46	**5.63**	1.23	1.590, p=.115
Thinking about Google/ Google Scholar						
I can work with this system to get the best results (SYS1)		**5.80**	1.24	**5.45**	1.60	-1.854, p=067
The system is good at finding information (SYS2)		5.56	1.37	5.32	1.45	
The system works well in suggesting information (SYS3)		5.55	1.52	5.24	1.58	
If asked, I could explain why these results were returned (SYS4)		5.30	1.77	5.16	1.66	
Thinking about finding this information						
It was easy to identify what the information was about (INF1)		**5.68**	1.49	**5.24**	1.55	-2.122, **p=.036**

It was easy to see why the information had been retrieved in response to my query	(INF2)	5.58	1.48	5.09	1.63	
I felt that my query had been understood	(INF3)	5.57	1.58	5.09	1.64	
It was easy to find my way round the information	(INF4)	5.56	1.47	5.11	1.54	
I got ideas for refining my search, when needed	(INF5)	4.77	2.10	4.71	1.78	

The final open question (Question 5.1, Appendix 1) asked participants to write down the 3 factors that were most influential in making their judgment on the usefulness of the information found. The written responses further suggest that differences in the formation of this judgment may be found across the Google and the Google Scholar tasks. Each suggested factor was categorised as referring to the '*credibility*' of the information, with mention of authority; the '*quality*' of the information, with mention of professional layout, structured, well written, accuracy; the '*relevance*', with mention of the query topic; and, '*scope*' of the information with mention of amount of detail, overview of topic, made sense or clarity of the information to me and enough information given. On Google, the 200 factors written down were categorised as referring to the credibility of the information (44%), quality of the information (11%), relevance (22%) and scope (23%). On Google Scholar the 198 factors provided were categorised as referring to the credibility of the information (40%), quality of the information (17%), relevance (28%) and scope (15%). Whilst the coding of these factors was checked for consistency amongst the researchers there is an element of subjectivity in the interpretation. However, with that caution, there is an indication that more reference to the quality of the information was given regarding the Google Scholar task (with statements on professional layout, structure, well written, accuracy) and some emphasis on the relevance of the information to the query. Whilst more reference was made to the scope of the information with regards to the information retrieved from Google for the Turing task (with statements on enough information given, the right amount of detail, overview of topic, clarity or easy to understand or made sense 'to me') and more mention of the importance of information credibility.

Factor Analysis

Factor analysis was used to extract the latent meaning underlying all the items, that is, the intended constructs when measured with multiple questions in the survey. If the statements (items) for a construct are well designed they should converge and form a major factor (Hair, Tatham, Anderson, & Black, 1998). Factor analysis with principal component analysis was used to determine the smallest number of factors to best represent the inter-relationships among the items in this study. Two separate analyses were conducted on each of the datasets (the Google task and the Google Scholar task) in order to explore differences between the factors that affected participants' formation of the judgment of usefulness. As recommended by Pallant (2007), Cronbach's alpha coefficient confirmed the reliability of the data in terms of internal consistency. In general, these ranged from 0.706 to 0.926 which is higher than the minimum cut-off, 0.70. The only exception was the possible construct of Scope with one item and with a low Cronbach's Alpha (0.69 Google and 0.45 GS) and so this was omitted from the further analysis. The Kaiser-Meyer-Olkin (KMO) values were greater than the recommended value, 0.6 indicating sampling adequacy of the data prior to conducting factor analysis. Table 3 reports the principal component analysis results with varimax rotation using SPSS. In extracting the major principal components as constructs and to satisfy convergent validity the factor loading was taken at greater than 0.5 (items below are indicated *). For consistency and to enhance comparability, when labelling we sought alignment across the factors for the Google and GS tasks although there are important differences in the composition of the comparable factors.

Table 3: Factors formed through principle components analysis for both data sets.

| Google Factors		Cronbach's alpha	Google Scholar Factors		Cronbach's alpha
G-CONTENT		0.87	GS-CONTENT		0.86
	Factor loading			Factor loading	
I can believe the content	.850		I think the content is accurate	.831	
The information appears to be authoritative	.847		The content is totally related to the topic of interest	.765	
The information is of good quality	.815		The content is very easy for me to understand	.753	
I think the content is accurate	.717		I can believe the content	.747	
G-RELEVANT		0.71	The information appears to be authoritative	.658	
	Factor loading		The information is well written	.652	
The content is very easy for me to understand	.837		The information tells me most of what I need to know	.613	
The content is totally related to the topic of interest	.549				
The information tells me most of what I need to know	.443*				

G-STYLE	0.79	GS-STYLE	0.72
	Factor loading		Factor loading
The information is well written	.847	The information is too detailed	.807
The author/org appears to be knowledgeable	.786	The author/org appears to be knowledgeable	.626
The information is clearly structured	.535	The information is of good quality	.517
G SCOPE	0.69*	GS-SCOPE	*0.45**
The subject area is only marginally related to the topic of interest		The coverage is too general for me	
The coverage is too general for me		The subject area is only marginally related to	
The info is too detailed		the topic of interest	
INFORMATION-EASE	0.85	INFORMATION-EASE	0.93
	Factor loading		Factor loading
It was easy to see why the information had been retrieved in response to my query	.868	I felt that my query had been understood	.906
It was easy to identify what the information was about	.837	It was easy to identify what the information was about	.873
It was easy to find my way round the information	.748	It was easy to find my way round the information	.868
I felt that my query had been understood	.719	It was easy to see why the info. had been retrieved in response to my query	.861
SYSTEM-EASE	0.86	SYSTEM-EASE	0.90
	Factor loading		Factor loading
The system works well in suggesting information	.893	The system works well in suggesting information to me	.864
The system is good at finding information that matches my query .	.882	The system is good at finding information that matches my query	.850
I can work with this system to get the best results	.829	I can work with this system to get the best results	.827
If asked, I think I could explain why these results were returned in response to my query	.651	If asked, I think I could explain why these results were returned	.823

In the Google Scholar data set, the factor labelled Content included the items relating to the credibility, authority and quality of the content as well as the items relating to the relevance of the information. This convergence of the items into a factor indicates that an overall impression of the information content and its relevance is a core holistic assessment of the information found. In comparison, the Google data set identified a possible distinct factor, G-Content formed on the items relating to the credibility, authority, quality and perceived accuracy of the information. In GS, relevance is embedded, but on Google relevance appears to be judged separately, or possibly not at all. It is notable that the items that were intended to reflect an assessment of the information as relevant did converge into a factor G-Relevant but that this has weak validity with a loading of < 0.5 for one of the items '*the information tells me most of what I want to know*' (.433*) leaving a factor of only two items. With that, only a weak factor of Relevance is indicated as a criterion in the evaluation of the information retrieved on Google. G-Content, however, is confirmed as a core factor. Further differences in the Content factors for Google and Google Scholar are shown in the items themselves. The first item (with the greatest factor loading) in the Google Content factor is *I can believe the content* whereas *I think the content is accurate* has the greatest loading in GS Content factor. Furthermore, whilst the item *The information is well written* is in the GS Content factor, in Google it forms the factor labelled Style with *The author appears to be knowledgeable* and *The information is clearly structured*. Similarly *The information is of good quality* appears in the Content factor in the Google data but forms the factor labelled Style in GS with the items *The author appears to be knowledgeable* and *The information is clearly structured*. This may suggest that participants are making some subtle distinctions in their perceptions of the Google and the GS information. Regarding the composition of the factors Information Ease and System Ease of Use the same items converged to these factors in both data sets with only a lower loading for the item *If asked, I could explain why these results were returned in response to my query* on the Google factor System-Ease and a higher factor loading for the item *I felt that my query had been understood* on the GS Information Ease factor.

Relationships Among the Factors

With the constructs measured with the multiple items in the survey the relationships between the independent (e.g., Content) and the dependent (i.e., Usefulness) can be explored. Possible predictors of the judgment 'usefulness of' and 'certainty found' were examined by running multiple regression analyses. If a factor is related to the judgment we can expect to obtain significant regression coefficients. In the Google data set the analysis was used to determine whether the factors G-Content and G-Style would successfully predict Usefulness (USE) and Certainty (CERT). G-Relevant was included in the analysis although noted as having a weak validity (and possibly related more to a perception of *ease in*

understanding the information rather than a relevance judgment). Figures 1a and 1b present the contribution of each predictor variable to the criterion variables. For USE the regression model obtained was statistically significant with the predictor variables G-Relevant, G-Content accounting for 53% of the variance in the criterion usefulness (USE) and, on certainty (CERT) G-Relevant and G-Content accounted for less than half (42%) of the variance at p<0.01. In both, G-Style was removed from the stepwise model as no significant correlation was found. Figures 1c and 1d present the contribution of each of the predictor variables Information and System Ease of use to the criteria Usefulness and Certainty. Information Ease resulted in a significant analysis of variance on USE (53%) and on CERT (54%). The System Ease construct did not predict Usefulness of the Google information but held moderate association at p<.005 (β= .208) with Certainty that appropriate information was found.
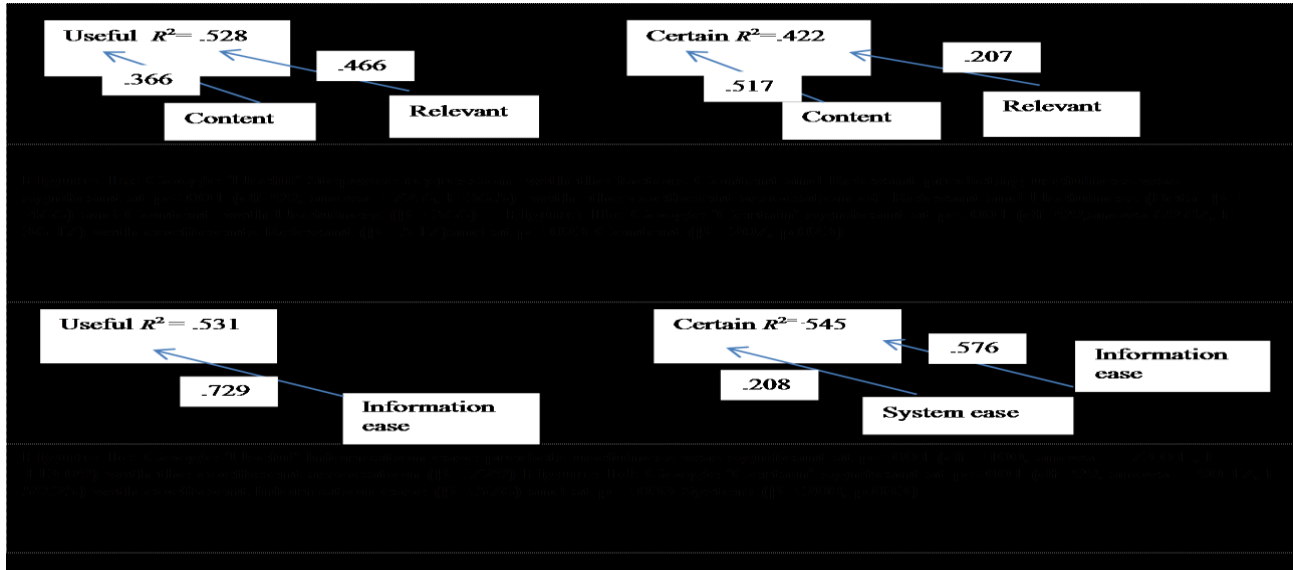
In Google Scholar the regression model on Usefulness and Certainty were predicted by the GS-Content construct (shown in Figures 2a and 2b) with a strong association and accounting for less than half 43% of the variance in the assessment of usefulness. GS-Content accounted for a lower 38% of the variance in Certainty judgment at P< 0.01 (Figures 2b) although again no correlation was found with GS-Style. When the model was built with the Information and the System Ease constructs both were found to predict Usefulness and Certainty (Figures 2c and 2d)

Figure 2a: GS 'Useful' Relevant–Content predicts usefulness was significant at p<.001 (df =100, anova = 66.52 , F 74.5) with the coefficient association (β=.653) **Figure 2b:** GS 'Certain' significant at p<.001 (df=100, anova = 58.73, F 61.6) with coefficient (β=.617)

Figure 2c: GS 'Useful' Information ease **and** system ease predicts usefulness was significant at p<.001 (df =99, anova = 72.48 , F 43.05) with the coefficient association information ease(β=.531) and system ease (β=.216) **Figure 2d:** GS 'Certain' significant at p<.001 (df=99, anova = 56.74, F 56.74) with coefficient Information ease (β=.455) and system ease (β=.212)

Further analysis was undertaken to examine the relationship of the factors of Information Ease and System Ease on the other judgments formed. Towards a measure of cognitive relevance, as distinct from usefulness, participants were also required to respond to items such as, *The information was just the right amount for me* and, *I expect to feel well informed on this topic* in an attempt to capture the users' forming a personal response to the information. In GS the further items loading to this factor included *The information has helped me to learn about the subject* and, *I will find this information easy to understand*. In Google the further items were *The information would help complete the assignment,* and *I expect to know more about the topic once I have read this information.* Again the convergence of the items for 'Cognitive Relevance' differ with what appears to be a more detached judgment made in the Google task. Furthermore, the multiple regressions when run on the Google and the GS Cognitive Relevance factors presented a different model of the influencing factors and their relations with USE and CERT. On the Cognitive Relevant factor both the Style and the System Ease of use factor accounted for a moderate amount of variance, especially on the GS data set. Specifically in the Google Scholar data set Information Ease and System Ease, as predictor variables, accounted for about half of the variance (45% in Cognitive Relevant) but, on Google, were low accounting for only 29% of the variance. The System Ease construct held a moderate association at $p < .001$ ($\beta = .457$) (Information Ease was lower at .292) and again a weaker association with Google Cognitive Relevant at $p = .007$ ($\beta = .331$).

Discussion

This study investigates the judgment of usefulness across task/engines to determine what influences this judgment in these different contexts and, specifically the factors that participate in determining this judgment. With regards to the searchers' evaluation of the information in task contexts and, more generally, the question 'what does it mean to be information literate in the specific context of web search?' this study highlights a difference in the evaluation of information when completing the Google and the Google Scholar task. Overall, the study has shown that the participants were making small but important differences in the evaluation of the information across the tasks. With regards to the task context, participants reported mid familiarity and knowledge, and a high certainty the information could be found, especially on Google. With regards to the view held on Google and GS, the highest means were obtained for the items *I can work the system*, *system is good* and w*orks well* indicating a positive view on the functionality of the system for the retrieval of information. On the descriptive statistics alone, differences were seen in the assessment of the information for the Google task and the GS task. Whilst these differences are subtle, Google information was judged on *accuracy*, *believeable* and *structured*, whereas GS information was judged on *accuracy*, *quality* and *well-written*. The notable difference is the emphasis placed on the credibility of the information and its quality.

Factor analysis further identifies and confirms the constructs of the evaluation of information as based on Content, Relevance, Style attributes as well as Information and System Ease of Use. Previous research on the criteria on which relevance judgments are made identify influencing factors such as style, accuracy and reliability (e.g., Schamber & Bateman, 1996) and criteria such as topicality and novelty (Xu & Chen, 2006). Further research into how people evaluate information when searching in specific contexts, such as for school coursework or for health information, identifies factors relating to design (Sillence, Briggs, Harris & Fishwick, 2007) or usability (Walraven, Brand-Gruwel & Boshuizeet, 2009) that contribute to the evaluation and selection of the information. The present study confirms that both the content factors, based on the attributes of the information and the factors relating to the ease in finding the information contribute in the evaluation of the information. Examination of the composition of these constructs again suggests differences in the evaluation of information for the Google and the GS tasks. The items with the greatest loading in the Google Content and Style factors suggest that Google information is judged as credible and good quality, as well as well-written. Whereas based on the loading of the items in the GS Content and Style factors, the information was judged as accurate, credible and well-written, as well as being of good quality. Again the difference pivots on the items relating to credibility and accuracy. The judgment of the Google information is associated with its perceived credibility, whereas perceived accuracy may have more influence with respect to the GS information. This was further indicated in the analysis of the factors with the items relating to structured information and knowledgeable author appearing, in the Google factor, with the item indicating the information is well written and in the GS factor with the item referring to 'good quality' These key differences in the evaluation of the information are shown in Figure 3 which suggest that the participants are making a more sophisticated judgment of the GS information as accurate and good quality compared to the less critical assessment of the Google information as believeable and well written based on authorship.
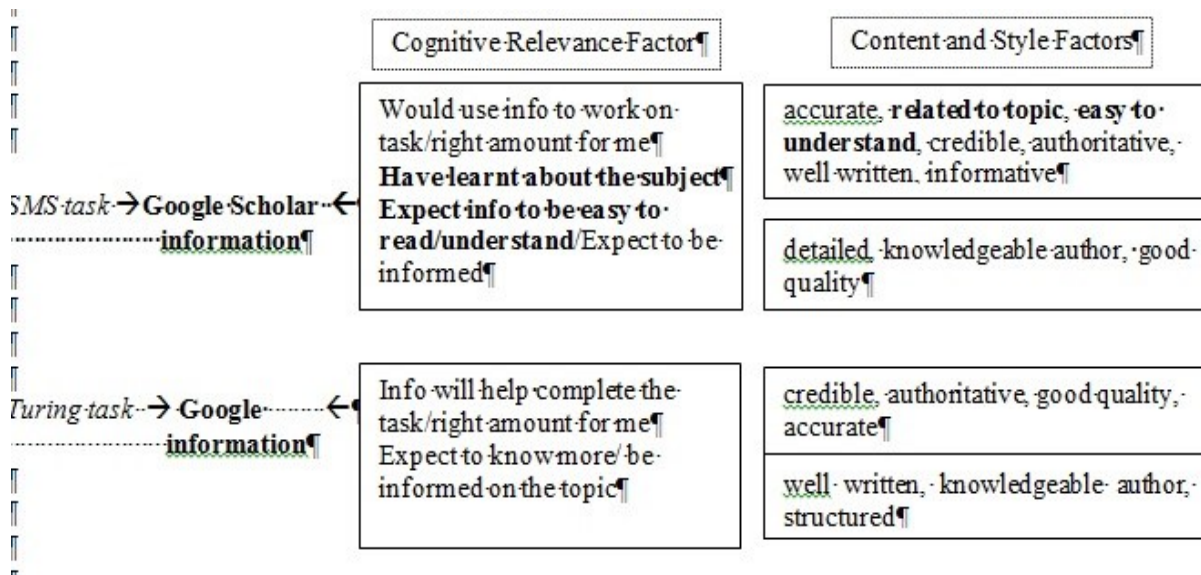
There is some evidence from previous studies that domain expertise affects evaluation of the credibility of online information where novices focus on the surface features of text, such as length, and, whereas the domain experts on the subject of the text will focus primarily on the semantic features, such as factual accuracy (Lucassen, T., Muilwijk, R., Noordzij, M.J. & Schraagen, J.M., 2013). In the present study, the effect of domain knowledge is not studied, however the participants' focus on the perceived quality and accuracy features predominantly in the evaluation of the GS information and less so with the Google information. Whether this is due to the difference in the type of information retrieved using Google and using Google Scholar or whether other factors, such as domain knowledge or, as it may appear in this study, the requirements of the task impacts on the users' evaluation of the information is partly addressed in the further analyses.

**Figure 3:** Key differences in the evaluation of the information.
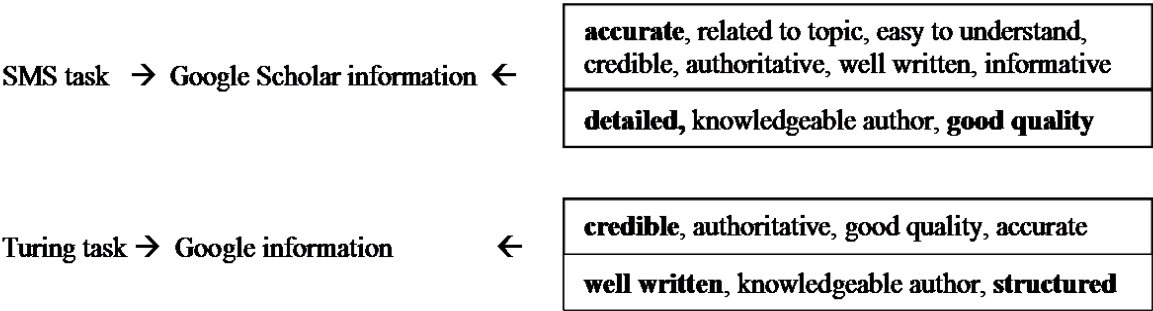
On the analysis of the descriptive statistics alone participants in the diary study appear to evaluate the information in forming a judgment of usefulness on slightly different criteria depending on the task and information retrieved. This is further evidenced through the strength of the association of the factors to the judgment of Usefulness. The final stage in the data analysis was to study the relation between the constructs as influencing factors in forming the judgments of the information usefulness, cognitive relevance and certainty. This provided insight into the judgments as depicted in Figure 4. On Google, the judgment is predicted or formed on the Content factor comprising the attributes of the information *believe the content*, *information is authoritative*, *good quality*, and *accurate* as well as weakly on some Relevance assessment (the factor comprising the items *related to the topic* and *easy to understand*). On GS the information was evaluated in terms of accuracy, relevance to topic, credibility and quality with the judgment of Usefulness predicted by the Content factor which, as was noted, contained items referring to relevance to topic as well as information accurate, believe and quality. Whilst this deserves further investigation, it does suggest that a different evaluation was conducted on the GS information with the embedding of the relevance judgment with consideration of quality, accuracy, credibility and relevance of the information, whereas the evaluation of the information for the Google task appears to have involved the user less in forming judgments of relevance.

This apparent difference in the judgment of the information across the tasks is explored further in the analysis of the items scored when asked to think about making use of the information (Question 6 in the questionnaire). The loading of these items to a factor labelled Cognitive Relevance attempts to capture the users' forming a personal response to the information, and examination of the factors formed, shown in Figure 4, furthers suggests that there are important differences in the evaluation of the information and the judgments across the two tasks. The indication that the participants were making an overall judgment of the relevancy of the content retrieved in the GS task was borne out in the analysis of Cognitive Relevance factors. In GS the items loading to this factor notably included*, The information has helped me to learn about the subject*, and *I will find this information easy to understand* and distinct from the items in the Google factor, *The information would help complete the assignment,* and *I expect to know more about the topic once I have read this information.* The items loading to Cognitive Relevance relate to a cognitive engagement with the information, although the different items in the factors formed across the data sets suggests that users were informed by the information retrieved on GS, with a greater emphasis on learning and an involvement with the information. The Google information appears to be assessed as useful, to be used at some later time, where perhaps the engagement with the information is to judge its potential, rather than actual, informativeness. Figure 4 thus builds on Figure 3 to illustrate the evaluation of the information and the difference in the judgments formed across the tasks.

Figure 4 Key differences in the judgments of the information

Cognitive Relevance Factor¶	Content and Style Factors¶

SMS-task➔Google Scholar ··· ←
··················· information¶

| Would use info to work on task/right amount for me¶ **Have learnt about the subject¶ Expect info to be easy to read/understand**/Expect to be informed¶ | accurate, **related to topic, easy to understand**, credible, authoritative, well written, informative¶ |
| | detailed, knowledgeable author, good quality¶ |

Turing task ··➔ Google ········· ←
················· information¶

| Info will help complete the task/right amount for me¶ Expect to know more/be informed on the topic¶ | credible, authoritative, good quality, accurate¶ |
| | well written, knowledgeable author, structured¶ |

Further understanding of the search is obtained through the analysis of the association of the 'contextual' factors, Information and System Ease as influencing the judgments formed. Across the tasks, perceived Information Ease, the apparent support in evaluating the information, associated with the assessment of Usefulness. Also, a significant difference was found in the higher ratings on the Information Ease item 'easy to identify what the information was about' on the information for the Google task, suggesting that ease in identifying what the information was about and why it is retrieved in response to the query also influences the judgment of the usefulness of the information, particularly in the Google task. Furthermore System Ease, the perceived effectiveness of the system, was found not to be an influencing factor in the Google task but was associated along with Content/Relevance and Information Ease in the judgment of Certainty that useful information had been found. It may be that the perceived System Ease, the effectiveness of the system, helps form a judgment of Certainty that useful information is found, but, in the Google task, does not necessarily influence the evaluation of the Usefulness of the information itself. This suggests that the look up task on Google (essentially, find some information on Turing) involved the searcher in making a surface judgment of usefulness influenced by a few core factors. The Google Scholar task, on the other hand, could be considered as more demanding with regards to the users' involvement with the search, requiring more consideration of the relevance of the information retrieved to the task to find information on texting and written English. System Ease, alongside Content and Information Ease, appeared as an influencing factor in the judgments of Certainty and Usefulness in the GS task. Furthermore the System Ease construct held a moderate association as a predictor of the judgment implied in the Cognitive Relevance factor, along with Style, on the GS task. The final Figure 5 depicts the environment of the information evaluation in the Google Scholar/SMS task and in the Google/Turing task contexts. Whilst this reflects the relationship between the core factors and the judgments formed of Usefulness and Cognitive Relevance, as this study was exploratory no attempt is made to model these findings with lines showing the association among the judgments formed and the core factors. The figure does however summarise the key finding of the different judgments and the influencing factors.

Figure 5: Information Interactions in two task contexts.
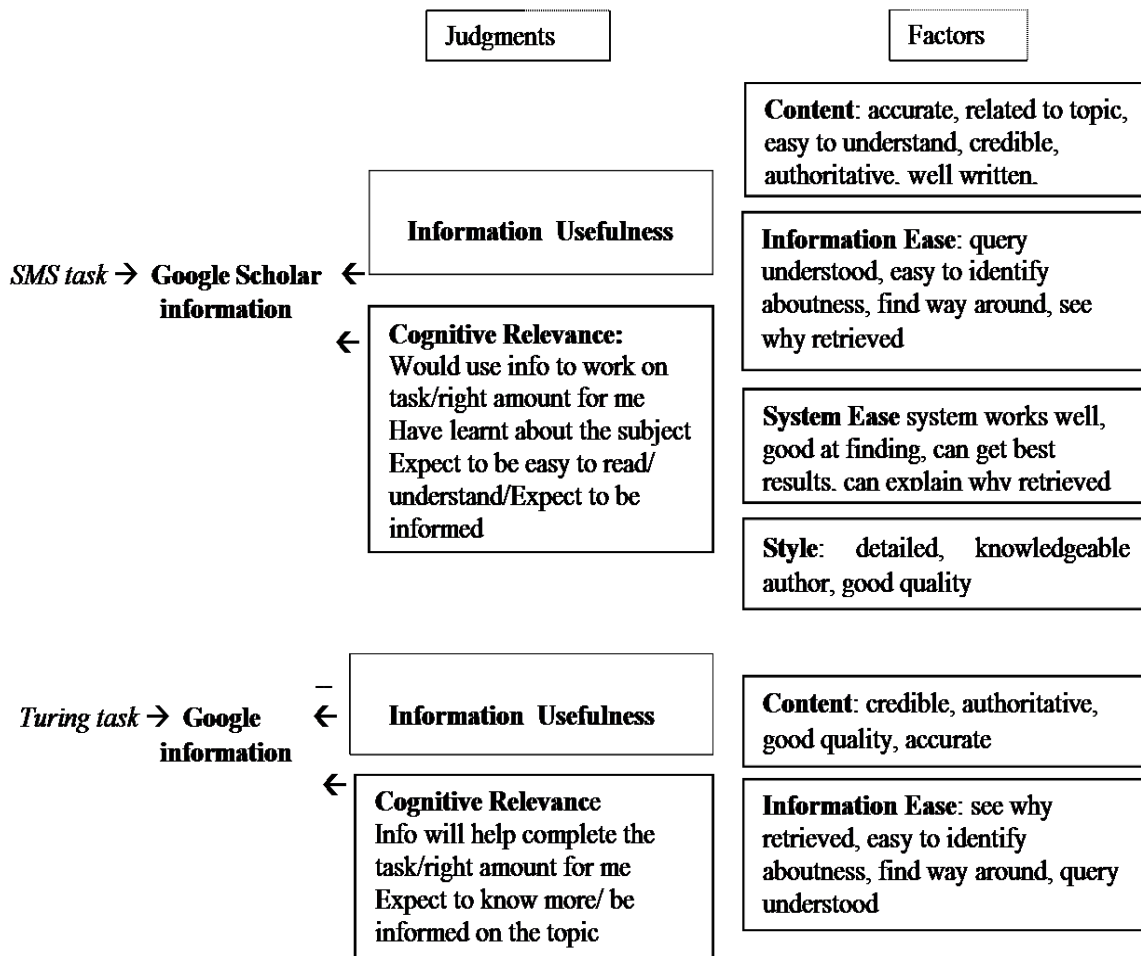
In this study, the user's perception of GS System Ease of Use (based on the responses to the items such as The system is good at finding information that matches my query) is shown, not as blind trust, (which we substitute for the more critical evaluation of the information when needed for the successful completion of the task) but rather, it is suggested as an influencing factor in the evaluation of the usefulness of the information. Perhaps the searcher is influenced by more and varied factors when the task demands an involved evaluation of the information, as with the GS task. Perhaps a positive view of the system facilitates the searcher in making relevance judgments? Further research is needed; however the results from this study highlighted the difference in the task contexts where perceived System Ease appears to influence the evaluation of the information when involving judgments of relevancy, but does not in the context where the evaluation is largely based on the credibility of the information and potential informativeness. Certainly the use of the information retrieved to form relevance judgments during a search places greater cognitive demands on the searcher and, in this study, results in the searcher drawing on a range of content and contextual factors.

The extent to which this study provides insight into what it means to be information literate in the context of web search is limited to the study's focus on the searchers' interaction with the information. The GS task involved the user in making relevancy judgments, including topical relevance and a type of cognitive relevance, which was characterised as the use of the information to learn about the subject and to form a personal perspective of relevancy. This interaction was not observed in the analysis of the Google data set where the Content factor, influencing the judgement of Usefulness, related to the credibility and other attributes of the information. Furthermore, when asked about the use of the information retrieved, participants focused on a rather fixed assessment of the potential rather than the actual informativeness of the information with respect to the search task. The active engagement of the user with the

information retrieved, particularly on Google Scholar, seems far removed from the view of the web searcher satisfied in finding something, anything and suggests that further research, focusing on the searchers' evaluation of information and the judgments formed, could lead towards a better understanding of the information interactions in search. The study of information searching tends to characterise the actions of search focusing on query attributes, search strategies and tactics (see for example, White, Dumais & Teevan, 2009). The study of search as a cognitive activity and interactions provides complementary insights into the process.

This investigation into the factors in the evaluation of information suggests a relation that has not previously been studied, that of the impact of the perceived Effectiveness and Ease of use of the search engine on the Evaluation of the information particularly when involved in relevance judgments. Previous research has mostly focused on the quality and relevance of the information as a key determinant of the user's evaluation of search engine effectiveness (for example, in Crudge & Johnson, 2004; Lewandowski & Höchstötter,2008) and many studies have focused on the effects of factors such as, the search task, web experience and information presentation on search behaviour and the evaluation of the system (Jansen, Spink & Pedersen, 2005; Lorigo et al., 2005). In particular, Nahl (1996) focusing on users' cognitive and affective states, found that self-efficacy and a confidence in one's ability to carry out a successful search influenced outcome expectations as well as performance. This research, instead, suggests a value in further study based on the impact of the perceived Effectiveness and Ease of use of the search engine on the evaluation of the information or on the cognitive activity (relevance judgements) of search. Whilst the approach to the investigation carried out here can only evince a relation between the users' Perception of the System Effectiveness and/or Ease of use and their involvement in, and evaluation of, the information it does suggest that perceived System Ease is an influencing factor in forming the critical judgments on the relevance of the information retrieved. It appears that the influence of the System Ease of use is prevalent in tasks that involve the user in interaction with the information as opposed to the task in which the user is making a more surface judgment on the information content.

Whilst this study does not present a model of information literacy in the context of web search, per se, it does presents a means to develop such a model through the study of searcher's critical evaluation of information. Further research could shed more light on the possibility that the searcher's positive perception of the system's effectiveness and/or ease of use influences the making of relevance judgments and/ or the converse that the perception of system ease of use has no relationship with the users' evaluation of the retrieved information usefulness particularly when no relevance judgments are formed. Perhaps the positive view of the system influences us to actively engage with and use the information retrieved to form relevance judgments and in the process make cognitive shifts on our understanding of the subject. Given that this is a complex interaction between the searcher and the information, it would seem reasonable that the view of the effectiveness of the system in retrieving the information has some role to play which is not otherwise observed in a straightforward evaluation of the credibility, quality and usefulness of the information found.

Conclusions

Information interactions in search contexts is typically studied by observation, particularly talk aloud for the depth of insights the method yields. The instrument (the questionnaire) employed in this pseudo diary study has allowed for a shift in the focus of attention from behaviour to the cognitive aspects in making the evaluative judgments of the usefulness of the information retrieved. The approach has confirmed that users (the student participants) form a judgment on the usefulness of the information retrieved for a task and that this is influenced by factors relating to the information content specifically its credibility and perceived quality. In particular, the questionnaire has helped explore the subtle differences in the constructs of the assessment of usefulness of the information retrieved for different tasks across the different engines. Considerable difference was found in the items relating to information content as influencing the evaluation of the information across the tasks, with a more involved judgment on the GS information than the seemingly simpler assessment of the information for the Google task. Furthermore, this study suggests that the perceived ease in which the information is found may also influence the judgment of the information usefulness and that it is more likely that the perceived system effectiveness forms part of the judgment of the information usefulness when this is a central activity in searching for information. These findings warrant further investigation to better understand the importance of the user perceived effectiveness and/or ease of use of the system on the types of critical judgments formed for the successful completion of the search task. Further refinement and application of the questionnaire could yield additional insight into the information interactions of search, focusing on the judgments formed and the associated influencing factors.

References

American Library Association (1989). Presidential Committee on Information Literacy. Final Report. Chicago: American Library Association.

Association of College & Research Libraries (ACRL) (n.d). Information Literacy Competency Standards for Higher Education. Retrieved from: http://www.ala.org/acrl/standards/informationliteracycompetency

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, 63(9), 1728–1746.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine. I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7), 646–665.

Blandford, A., & Attfield, S. (2010). *Interacting with Information,* Synthesis Lectures on Human-centered Informatics. Morgan & Claypool, San Rafael, CA, USA.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913–925.

Brown, G. (2001). Locating categories & sources of information: How skilled are New Zealand children? *School Library Media Research*, 4, 28–33.

Coiro, J. (2003). Reading comprehension on the Internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher*, 56(5), 458–464.

Crudge, S.E., & Johnson, F.C. (2004). Using the information seeker to elicit construct models for search engine evaluation. *Journal of the American Society for Information Science and Technology*, 55(9), 794–806.

DeVellis, R F. (2003). *Scale development: Theory and application.* (2nd ed.) Thousand Oaks, California: Sage.

Fosmire, M., & Macklin, A. (2002). Riding the active learning wave: Problem based learning as a catalyst for creating faculty-librarian instructional partnerships. *Issues in Science and Technology Librarianship,* 34. Retrieved from: http://www.istl.org/02-spring/index.html

Hair, J.F., Tatham, R.L., Anderson, R.E., & Black, W. (1998). *Multivariate Data Analysis.* (5th ed.) Prentice Hall, Upper Saddle River, NJ, USA.

Hargittai, E., Fullerton, L., Menchen-Trevino, E. & Thomas, K.Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4, 468–494.

Helms-Park, R., Radia, P., & Stapleton, P. (2006). A preliminary assessment of Google Scholar as a source of EAP students' research materials. *Internet and Higher Education*, 10(1), 65–76.

Jansen, B.J., Spink, A., & Pedersen, J. (2005). A temporal comparison of Alta Vista Web searching. *Journal of the American Society for Information Science and Technology,* 56(6), 559–570.

Jansen, B.J., Zhang, M., & Schultz, C.D. (2008). The effect of brand on the evaluation of IT system performance. *Proceedings of the 11th Southern Association for Information Systems Conference (SAIS 2008)*, March 13–15th, Richmond, USA.

Kuhlthau, C.C., Heinstrom, J., & Todd R.J. (2008). The 'information search process' revisited: is the model still useful? *Information Research,* 13(4). Retrieved from: http://www.informationr.net/ir/13-4/paper355.html

Kelton, K., Fleischman, K.R., & Wallace, W,A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59 (3), 363-374.

Leu, D. (2000). Literacy and technology: Deictic consequences for literacy education in an information age. In M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds.), Handbook of reading research (pp. 743–770). Mahwah, NJ: Erlbaum.

Lewandowski, D., & Höchstötter, N. (2008). Web Searching: A Quality Measurement Perspective. In A. Spink & M. Zimmer (Eds.), Web Search: Multidisciplinary perspectives, (pp. 309–340). Springer, Berlin.

Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2005). The influence of task and gender on search and evaluation behavior using Google. *Information Processing and Management,* 42(4), 1123–1131.

Lucassen, T., Muilwijk, R., Noordzij, M.J., & Schraagen, J.M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science & Technology,* 64(2), 254-263.

Mullen, L.B., & Hartman, K.A. (2006). Google Scholar and the library web site: The early response by ARL libraries. *College & Research Libraries,* 67(2), 106−122.

Nahl, D. (1996). Affective monitoring of Internet learners: Perceived self-efficacy and success. In S. Hardin (Ed.) Proceedings of the 59th Annual Meeting-American Society for Information Science, Vol. 33 (pp. 100–109). October 20–25th, Baltimore, MD.

Pallant, J. (2007). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows.* (3rd ed), Open University Press, Milton Keynes, UK.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G. & Granka, L. (2007). In Google we trust: users' decisions on rank, position and relevance. *Journal of Computer-Mediated Communication,* 12(3), 801–823.

Pritchard, F., & Cartwright, F. (2004). Transforming what they read: Helping eleven-year olds engage with Internet information. *Literacy,* 38(1), 26–31.

Rieh, S., & Hilligoss, B. (2008). College students' credibility judgments in the information seeking process. In M.J. Metzger (Ed.), Digital Media, Youth and Credibility (pp. 49–71). Cambridge, MA: MIT Press.

Sampanes, A.C., Snyder, M., Rampold-Hnilo, L., & White, B.K. (2011). Photo diaries – A peek into a mobile worker's life. In A. Marcus (Ed.), Design, user experience and usability. Theory, Methods, Tools and Practice. Lecture notes in Computer Science, (pp. 640–647), Vol. 6769.

Saracevic, T. (1975). Relevance: A review and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science and Technology*, 26(6), 321–343.

Schamber, L., & Bateman, J. (1996). User criteria in relevance evaluation: Towards development of a measurement scale. *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Vol. 33 (pp. 218–225), Baltimore, MD, October 20–25.

Sillence, E., Briggs, P., Harris, P. & Fishwick, L. (2007). How do patients evaluate and make use of online health information? *Social Science & Medicine*, 64(9), 1853–1862.

Tang, R., Shaw Jr., W.M. & Vevea, J.L. (1999). Towards the identification of the optimal number of relevance categories. *Journal of the American Society for Information Science and Technology*, 50(2-3), 254–264.

Toms, E., Freund, L., Kopak, R., & Bartlett, J.C. (2003). The effect of task domain on search. *CASCON, 3 Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, (pp. 303–312). IBM Press, Toronto.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, 52(1), 234–246.

Wang, P., & Soergel, D.A. (1998). A cognitive model of document use during a research project: study 1. Document selection. *Journal of the American Society for Information Science and Technology*, 49(2), 115–133.

White, R.W., Dumais, S.T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behaviour (pp. 132-141). Proceedings of *WSDM, February 9-12, Barcelona, Spain*.

Xu, Y., & Chen, Z. (2006). Relevance judgment: What do information consumers consider beyond topicality? *Journal of the American Society for Information Science and Technology,* 57(7), 961–973.

Appendix 1: The Questionnaire

What you are asked to do:
Please use *Google* and *Google Scholar* for these search tasks (see below). Spend as long (or short) as you would normally do when searching for information. We want you to look for some information to your best ability and stop either when you are satisfied with the information found/consider it is the best that you can find, or you have spent far too long on the task (we suggest that you spend **max 20 minutes** on each task).

You are asked to complete a questionnaire following each task. Once you have found some information you will be asked to choose up to 3 items and judge the usefulness of this information. It is very important that you have spent some time reading and thinking about this information as you are asked to form an evaluative judgement of its use for the task set. Therefore, you should keep in mind that this is not a test of your ability to search for information, your aim is to find information that you may decide to use (or not to use) to complete the task.

Tasks:

1. You are writing an assignment on young people's use of SMS and its effect on their written language. Use **Google Scholar** to find information to help you write this assignment.

2. Use **Google** to find information to help prepare an assignment on Alan Turing's contribution to the development of Computing.

You have two questionnaires, one for each task. Please tick which task this questionnaire refers to:

○ *Find information to help you write an assignment on the effect of young people's use of SMS.*

○ *Find information to help prepare an assignment on Alan Turing's contribution to Computing.*

1. Please complete the following before starting the search:

(**7**= very;…...**1**= not at all)

	7	6	5	4	3	2	1
How certain are you that the information could be found?	○	○	○	○	○	○	○
How certain are you that you will find the info?	○	○	○	○	○	○	○
I am used to searching for this type of query.	○	○	○	○	○	○	○
How familiar are you with the task topic?	○	○	○	○	○	○	○
How knowledgeable are you about this topic?	○	○	○	○	○	○	○
If asked, how willing would you be to inform others about this topic?	○	○	○	○	○	○	○

2. About the search:
2.1. Please record your query as input to the search box:

2.2. Were you content with the information you found *after your initial query*?   (**7**= very;…...**1**= not at all)

7	6	5	4	3	2	1
○	○	○	○	○	○	○

If not, please record your revised query below:

2.3. Please provide the url and/or author/title of at least one and up to 3 sources of information that you have read with respect to the set task (*remember that we are asking you to judge the usefulness of this information*):

> 1.
> 2.
> 3.

3. Thinking about the information just found:

(**7**= very;…...**1**= not at all)

	7	6	5	4	3	2	1
Please assess the usefulness of the information found with regards to addressing the task.	○	○	○	○	○	○	○
How certain are you that appropriate information has been found?	○	○	○	○	○	○	○

4. Thinking about finding this information:

0= I have not considered this at all   (**7**= strongly agree;……**1**= strongly disagree)

	0		7	6	5	4	3	2	1
It was easy to identify what the information was about.	○		○	○	○	○	○	○	○
It was easy to see why the information had been retrieved in response to my query.	○		○	○	○	○	○	○	○
I felt that my query had been understood.	○		○	○	○	○	○	○	○
I got ideas for refining my search, when needed.	○		○	○	○	○	○	○	○
It was easy to find my way round the information.	○		○	○	○	○	○	○	○

5. Thinking about the information you chose as being the most useful for the task:

0= I have not considered this at all   (**7**= very much so;……**1**= not at all)

	0		7	6	5	4	3	2	1
I can believe the content.	○		○	○	○	○	○	○	○
I think the content is accurate.	○		○	○	○	○	○	○	○
The information appears to be authoritative.	○		○	○	○	○	○	○	○
The information is of good quality.	○		○	○	○	○	○	○	○
The content is totally related to the topic of interest.	○		○	○	○	○	○	○	○
This subject area is only marginally related to the topic of interest.	○		○	○	○	○	○	○	○
The information tells me most of what I need to know.	○		○	○	○	○	○	○	○
The information is too detailed.	○		○	○	○	○	○	○	○

	0		7	6	5	4	3	2	1
The coverage is too general for me.	○		○	○	○	○	○	○	○
The content is very easy for me to understand.	○		○	○	○	○	○	○	○
The author/org appears to be knowledgeable.	○		○	○	○	○	○	○	○
The information is clearly structured.	○		○	○	○	○	○	○	○
The information is well written.	○		○	○	○	○	○	○	○

5.1. Please write down the **3 factors** which were most influential in making your judgment of the usefulness of the information:

> 1.
> 2.
> 3.

6. Thinking about making use of this information:

0= I have not considered this at all   (**7**= strongly agree;……**1**= strongly disagree)

	0		7	6	5	4	3	2	1
I would use this information to work on the task in hand.	○		○	○	○	○	○	○	○
If asked, I think that other people would value this information.	○		○	○	○	○	○	○	○
If I knew someone was looking for this information, I would tell them about this.	○		○	○	○	○	○	○	○
I think I will find the information easy to understand.	○		○	○	○	○	○	○	○
The information would help me to complete the assignment.	○		○	○	○	○	○	○	○
This information has helped me to learn about the subject.	○		○	○	○	○	○	○	○
I expect to know more about the topic once I have read this information.	○		○	○	○	○	○	○	○
I expect to feel well informed on this topic.	○		○	○	○	○	○	○	○
The information found seems to be the right amount for me.	○		○	○	○	○	○	○	○

7. Thinking about Google/Google Scholar (*which one was used for the task*):

0= I have not considered this at all   (**7**= strongly agree;……**1**= strongly disagree)

	0		7	6	5	4	3	2	1
I can work with this system to get the best results for my query.	○		○	○	○	○	○	○	○
If asked, I think I could explain why these results were returned in response to my query.	○		○	○	○	○	○	○	○
The system is good at finding information that matches my query.	○		○	○	○	○	○	○	○
The system works well in suggesting information to me.	○		○	○	○	○	○	○	○

8. Please tell us a little about yourself:

Your gender: ○ Male   ○ Female

Your age: [   ]

Course level: ○ UG Yr1   ○ UG Yr2   ○ UG Yr3   ○ Postgraduate

Course subject: [   ]

Student status: ○ UK   ○ International