

An investigation into Scottish teenagers' information literacy and search skills.

David Brazier¹, Geoff Walton² and Morgan Harvey¹

¹ Computer and Information Sciences Department, Northumbria University, Newcastle upon Tyne, UK

² Department of Languages, Information and Communications, Manchester Metropolitan University, Manchester, UK

Abstract

Introduction. *This paper presents the results of a study investigating the information literacy and search skills of young people in Scotland.*

Method. *Participants (n=57) completed two out of four different search tasks for which the correct answers (i.e. relevant documents) were known. Their interactions with the search system were logged and information about their own perceptions of the task were collected through pre- and post-task questionnaires.*

Analysis. *The log data from the search system was analysed using the R statistical software package to understand the performance and behaviour of the participants when conducting the search tasks.*

Results. *While we identified some evidence that information literacy and search skills were being employed, overall performance was low with participants often unable to produce successful queries and/or unable to identify relevant documents, even when some were present in the results.*

Conclusions. *Search performance and ability to identify relevant information was generally poor, a fact that participants themselves were frequently unable to recognise. The results also suggest a reliance on complex search assistance tools (such as spell checking and query suggestions), which are common features of major search engines, but not of smaller systems, which pupils are also likely to have to use.*

Introduction

The use of web search is now ubiquitous in daily life, whether from a work, study or social perspective, it is a constant aid in the quest for everyday information seeking. Many governments and local authorities increasingly offer their services, sometimes exclusively, through online means (Helbig, 2009). While this may lead to several benefits, there is concern about the expectation this places on people's information literacy and search skills. Is the use and reliance on assistive functionality luring users into a false sense of security and lessening the likelihood of the user employing information literacy skills to build good queries and assess the documents returned? And do information providers assume too much ability on the part of their end-users to search for, understand, evaluate and synthesise the information provided. Information literacy is a lifelong learning process that goes about providing the skills, knowledge and understanding required to ensure informed decision making and effective problem solving (Shenton and Pickard, 2014). These abilities are so crucial that many governmental organisations and politicians assert that they are essential to be able to participate effectively in our modern information society (Sturges, 2010). Within education, such skills are perhaps even more crucial, as pupils are often given tasks that necessitate these abilities and are first introduced to these concepts in school. To write a report one must identify the information required, use effective search strategies to locate relevant documents, evaluate the quality and

veracity of these and, finally, synthesise the information by converting it into a coherent narrative (Eisenberg, 1990).

Unfortunately, evidence suggests that *millennials* are not as information literate as one might expect and that the steps taken in education systems are not ensuring school leavers are equipped with these valuable skills (Pickard, et.al. 2013; Pickard et.al. 2014). Research shows that pre-university students are unable to construct effective search queries (Harrop, 2012), rarely employ effective search strategies (Coiro and Dobler, 2007) and are easily discouraged when a search engine does not immediately return useful results and feel swamped by non-relevant and poor-quality material, leading them to often simply cite/copy top-ranked resources without assessing their quality (Smith and Hepworth, 2007).

This is a problem which needs to be countered early on in a child's development as recent research questions whether student information behaviours can be changed once they enter Higher Education (Rowlands and Nicholas, 2008). Therefore, it is crucial that pupils be taught these skills early in their educational lives and be encouraged to apply them throughout.

In this work we investigate the information literacy and search skills of young people at an important stage in their education – shortly before they embark on their first set of major examinations. Unlike previous work, we evaluate this primarily based on quantitative data obtained from a large-scale user study (n=57) with participants from a large secondary school in greater Edinburgh, Scotland. We follow a well-tested methodology from the field of Information Retrieval to ensure that the studies are repeatable, representative and, crucially, in-context.

Related work

There has been considerable research into the search behaviours of children over the past twenty years, with studies focusing on children's search behaviours whilst using the web directory Yahoologans! (Bilal, 2000; 2002), adoption of search roles during the search process (Druin, 2009), behaviours extracted from large scale query logs (Duarte Torres et al. 2010; Duarte Torres and Weber, 2011), Internet searching on complex problems (Schacter et al., 1998), and adolescent search roles (Foss et al., 2009).

Bilal's contribution to the investigation into children's search behaviours has been extensive, with several works focused on children's interactions with Yahoologans when conducting various task types (2000; 2003). In these studies, a mixed methods approach collected quantitative data on the interactions with and traversal of the search interface while post study interviews were used to collect qualitative data on student's task generation, task preference and search interface opinion. These studies found that they preferred completing their own self-generated tasks, were more successful when browsing then when using keyword search and that the search interface design was confusing to most, and were a factor in the search breakdown experienced by the children. Although seminal in their contribution, the study data were collected in 1998, a limitation noted by the author in these papers a few years later.

More recently, Bilal investigated the reading level of Google's search results (2013) by comparing them to the flesch readability formulae, finding that a high number of results retrieved by Google were difficult or confusing to children attending middle school in the US. This paper compliments Bilal's findings by describing the effects that reading snippets, a short description of the documents content, has on the number of documents bookmarked and its effects on performance.

Duarte Torres and colleagues have observed large scale query logs to extract children's search behaviours and tested for rational sentiment in their search queries through the presence or absence of sentiment in the terms utilised (2010; 2011). Owing to the large scale of the studies and the numbers of participants, this approach is useful in providing generalizable results, however, it lacks specific context and makes some (well-reasoned) assumptions as a result of eliciting sentiment. Although understandable given the nature of the data collection, this paper

builds on this work by focusing on more specific contextual effects on query formulation and semantic notion towards the tasks performed.

Children are able to identify information overload as a potential problem of Internet search and have the ability to self-identify their skills shortage when it comes to refining search terms, which in turn leads to frustration, despite receiving information skills training, albeit *in isolation and not related to topic work* (Smith and Hepworth, 2007, p. 9). This paper seeks to further investigate these areas, through participant responses reflecting on these points as part of the pre and post questionnaire.

Coiro and Doblers' paper on reading comprehension of online texts identified that when given a typical school-based task that is, *to locate, evaluate and synthesize content area information within informational websites and search engines* (2007, p. 221), prior knowledge, inferential reasoning strategies and self-regulated reading processes were required by students to ensure successful online reading. Rowlands et al. (2008) identified that children under thirteen are *unable to construct effective searches and evaluate the results* (p.22), largely attributable to knowledge deficiency of domain specific information, a lack of understanding of how search engines work, difficulties in switching between the use of natural language and search queries, and a limited command of vocabulary to utilise synonyms for query reformulation. Also mentioning that young people have difficulty in selecting sufficient search terms, fail to evaluate information from an electronic source, and declined to undertake additional searches if information was already found, (Rowlands et al., 2008). Rowlands goes on to state the need for more research around young people's research information behaviours. This paper builds on these foundations by investigating the information literacy skills of Scottish secondary school students.

Methodology and data collection

The study employs quantitative analysis of secondary school-aged (ages 13-14) children's search behaviour through analysis of logs from a series of 30-minute information retrieval tasks, triangulated with qualitative assessments of the participants' own behaviour. This methodology is common in the field of information retrieval as it provides an unbiased and repeatable environment in which performance and behaviour can be exhaustively logged and evaluated. Furthermore, rather than simply asking students to abstractly comment on their information literacy skills, this study gives participants an in-context work task to perform, allowing us to evaluate their responses to this in a more naturalistic fashion.

We obtained permission to run our studies with pupils in year 3 (ages 13-14) at a large mixed-intake secondary school in the city of Edinburgh, Scotland. More details about the participants and recruitment are given later in this paper. Each pupil was given a desktop computer and asked to complete two information retrieval tasks (or '*topics*') randomly selected from a set of four. The tasks were taken from the 2005 TREC HARD collection, which provides a complete set of 100 topics, each of which is presented as a title and short description of the information need. The school's librarian and a selection of teachers were asked to reduce these to four topics based on topic suitability and perceived interest for the target demographic. The final four topics selected were:

- 347 - wildlife extinction;
- 353 - Antarctica exploration;
- 367 - piracy;
- 408 - tropical storms.

Pupils used a bespoke search system to collect a small set of relevant documents for each of their two tasks over a time-constrained period of between fifteen minutes and half an hour. The search system was designed to be similar in design to those the pupils are likely to be familiar with (e.g. Google search) and is therefore composed of the standard search bar and button with results presented as a list of 10 blue links with associated *snippets* of text (see *Figure 1*). The

entire user study was conducted using the search system, which logged all of the pupils' interactions, including queries entered, documents read and documents selected (*bookmarked*).

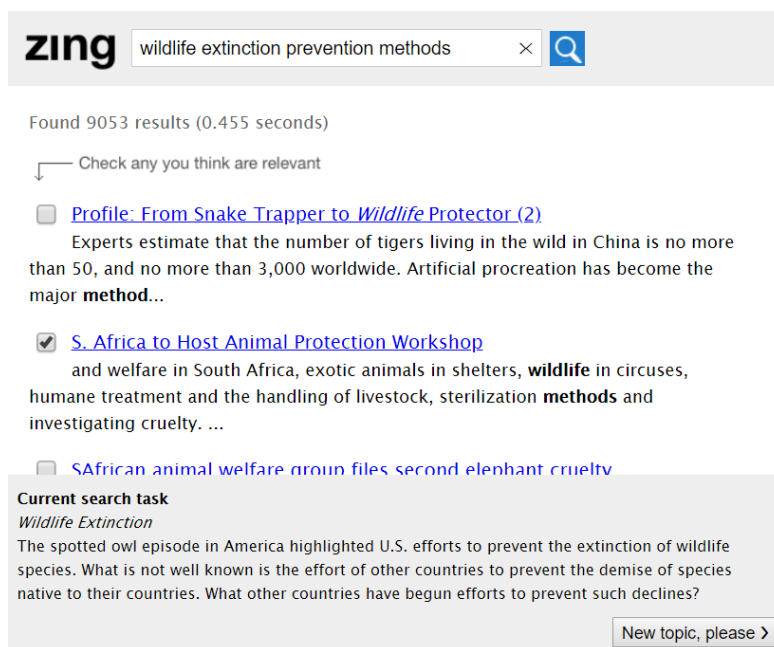


Figure 1: The *zing* search interface

The documents were taken from the TREC AQUAINT collection, a set of over a million documents from three large news agencies collected between 1996 and 2000. This is a complete information retrieval evaluation collection, meaning that there are pre-defined search topics associated with it together with relevance judgements for each of these topics. Relevance judgements are per-topic evaluations indicating which documents in the collection are relevant, and which aren't. Such a collection can be used to evaluate the performance of a new retrieval system or, crucially in this case, the search performance of users.

Participants filled in pre- and post-task questionnaires in which they self-assessed various elements related to search. Pre-task questions included those related to familiarity with and interest in the topics and expected difficulty in completing the tasks. Post-task questions focused on perceived learning and search success and trust in the retrieved documents. For each topic, the participants were asked to imagine that they'd be writing a report about it soon and needed to collect some relevant documents to help them in doing so. Documents in search results could be *bookmarked* (i.e. indicated as relevant) by means of checkboxes.

To obtain additional qualitative data from participants about their search behaviour and strategies, at the end of each session four participants, who all had one topic in common, were chosen at random by the system and were each asked by the research a question about their earlier search behaviour. Participants were shown either: their own search queries for one of the topics they had been given and asked how they came up with the search terms; or a document they had bookmarked and asked what made them choose that specific document. The assignment of questions to participants and a recording of the comments they each made was done by a bespoke web application, which ran on an iPad device.

The user study was piloted in October 2017 by a small number of participants (n=12) before proceeding with the full experiments in November and December 2017. Studies were performed with groups of up to fifteen students at a time over five separate sessions in one of the school's dedicated IT rooms.

Measures and metrics

Performance can be determined in a number of ways:

The quality of the participants' queries can be assessed by means of the total number of *hits*, that is the number of relevant documents appearing within the 10 search results, and the average precision of the search results. Precision is defined as simply the number of relevant documents returned divided by the count of all documents returned (i.e. both relevant and non-relevant). This is calculated for all of the positions in the search ranking (positions 1 through 10) and averaged to produce the *average precision*.

The ability of the participants to identify relevant documents can be determined by the number of relevant documents they managed to bookmark (select) per topic and, like the precision above, the ratio of relevant documents bookmarked over the total number bookmarked for each topic.

All pre- and post-task questionnaire responses are based on a 5-point Likert scale in ascending order of agreement from 1 (disagree) to 5 (agree).

Analysis

Quantitative data analysis consisted of descriptive statistics, Wilcoxon signed rank testing and linear model regression to examine the relationships between the research variables using the statistical software package R.

Findings

Study participants

Participants were recruited by the secondary school's librarian who asked all of the pupils in year 3 (total=; aged between 13 and 14) to participate. A consent form was given to each pupil to take home to be signed by their legal guardian before any involvement in the study could take place. In total, 60 pupils were able to take part in the experiments, however, some data loss occurred during the transition process from data collection to data cleaning, three users' search data were either unavailable or incomplete resulting in a final study population of fifty-seven (n=57).

The only demographic data recorded for each participant was their Scottish Index of Multiple Deprivation (SIMD) vigintile ranking. The SIMD is a '*tool for identifying areas of poverty and inequality across Scotland*' (Scottish Government, 2016, p. 2) and considers multiple indicators of deprivation (Income; Employment; Education; Health; Access to Services; Crime and Housing) which are grouped together. Areas of Scotland are split into data zones with roughly 760 people per zone. The data zones were then split into 20 groups (vigintiles), each containing 5% of Scotland's data zones. Vigintile 1 contains the 5% most deprived data zones, vigintile 2 the second most deprived 5% and so on, until vigintile 20 contains the least deprived 5%. This facilitates the identification of deprived areas, although it must be noted that '*not all deprived people live in deprived areas*' and '*not everyone within a deprived area is deprived*' (Scottish Government, 2016, p. 7).

The average SIMD ranking for the study participants was 17.5, with a range of 18, putting this group within the 87.5 percentile of Scotland on average or within the range of 10% of the most deprived and 5% of the least deprived areas. The participants were divided into groups based on the SIMD ranking by quartiles, with those in the first quartile scoring a SIMD ranking of nine or less, and those in the third quartile scoring nineteen or above. The aim of this grouping being to facilitate comparison across deprivation areas by performance and behavioural metrics.

Pre-task knowledge, interest and perceptions of task

Participants generally reported a low level of pre-study understanding of the topics - 2.35 on average. Interest in most of the topics was reasonably high at 2.94, suggesting that the topics selected were appropriate for the age group and of at least some interest. This is important as a lack of interest tends to result in a lack of engagement, therefore we can expect the relatively high levels of interest here to translate into a greater likelihood that pupils will engage fully with the search tasks. The level of clarity on how to complete the task was high at 3.85, indicating that most participants understood what they were being asked to do. Perceptions of how difficult it would be to find relevant documents and the overall task difficulty were 2.71 and 2.88 respectively, while the difficulty in knowing when to decide enough information had been found was reasonably high at 3.14.

The highest prior knowledge was for topic 347 (wildlife extinction; mean=2.88) and least for topic 353 (Antarctica exploration; mean=1.96). Interest was also reasonably high for three out of the four tasks, ranging from 3 to 3.38, although there was considerably less interest in topic 367 (piracy; mean=2.41). Participants noted it was quite clear what was required to complete the task for each topic with 353 the highest at 4.000 and 347 the least at 3.625. Judging whether they thought the task would be difficult, participants noted that topic 367 would be reasonably difficult at 3.2 with the other three topics approximately moderate in difficulty. Finding relevant documents was expected to be moderately difficult for all topics with topic 367 at 2.862 and 353 a close second at 2.833.

Querying performance

Overall performance was low with a median average precision of just 0.008 and a mean of 0.05. In terms of number of hits, the median number was only 1 and, out of a total of 600 queries, more than a third (227) returned not a single relevant result.

By topic (see Table 1), performance was low across all tasks, but particularly for topic 353 for which the median number of hits was 0. The reasons for this are discussed later.

Topic	Average Precision			Hits		
	Median	Mean	SD	Median	Mean	SD
347 – Wildlife extinction	0.05	0.08	0.69	3	2.54	1.87
353 – Antarctica exploration	0.00	0.009	0.02	0	0.38	0.73
367 – Piracy	0.02	0.07	0.1	1	2	2.36
408 – Tropical storms	0.02	0.06	0.07	1	2.12	1.92

Table 1: Average precision by topic

Previous research has suggested that young people tend to over-estimate their information seeking and literacy skills (Pickard, 2004). As Bilal writes: ‘*Although all children admitted an understanding of the search task...most of them were still unclear as to the type of information sought*’ (Bilal, 2000). It is clear from the pre-task questionnaire results that our participants were also quite confident in their ability to complete the tasks, however, this confidence is not really warranted, given their poor querying performance.

Task fatigue was anecdotally observed by the researcher after just one task, we therefore compare the average precision of each topic by the order in which they were completed (see Table 2). There was slightly better performance in sequence 1 topics, however differences are marginal, and still very poor overall, mean = 0.054 (sequence one) vs. 0.045 (sequence two). These differences are not significant ($p=0.1319$), although it is notable that the most difficult topic (353) displays the largest drop in performance.

Topic	Seq 1	Seq 2
347 – Wildlife extinction	0.090	0.073
353 – Antarctica exploration	0.012	0.005
367 – Piracy	0.080	0.057
408 – Tropical storms	0.053	0.060

Table 2: Average precision by sequence per topic

By topic, users performed best in topic 347 and worst in 353 in both the first and second sequence, and with the exception of topic 408, which had an increase, there was a decline in average precision in the second sequence. These differences were not significant (p -value = 0.1319).

Performance between SIMD quartiles is approximately equal with the median average precision at 0.01 for both quartiles and the mean 0.045 and 0.052 for the first and fourth quartiles respectively ($p = 0.998$). The mean number of hits was 1.61 and 1.72 for the first and fourth quartiles. When performance across topics is compared (see below), there are some interesting, albeit not significant, differences. Performance by the fourth quartile is much higher for topic 347 than for those in the first quartile and is higher than in any other topic.

Topic	Q1	Q4
347 – Wildlife extinction	0.03	0.105
353 – Antarctica exploration	0.011	0.013
367 – Piracy	0.062	0.08
408 – Tropical storms	0.064	0.058

Table 3: Average precision by quartiles across topics

These analyses suggest that the pupils from the fourth quartile SIMD band (i.e. those likely to be from wealthier backgrounds) are able to perform slightly better than those from the lowest quartile, especially when the topic is well known and interesting.

Bookmarks and reading

When determining performance, we can also consider per query how many relevant documents the participants were able to identify (bookmark) and how many they read. In general, participants did not manage to bookmark or read very many documents – on average they only bookmarked 0.97 and read 1.6 documents per query. Of those that they did choose to bookmark or read, only 36.3 and 23 percent were actually relevant.

Topic	Total number		Of which relevant (ratio)	
	bookmarked	read	bookmarked	Read
347 – Wildlife extinction	1.78	2.02	0.54 (0.3)	0.69 (0.34)
353 – Antarctica exploration	0.64	1.08	0.07 (0.11)	0.07 (0.07)
367 – Piracy	0.84	1.54	0.53 (0.63)	0.49 (0.32)
408 – Tropical storms	1.11	2.02	0.38 (0.43)	0.42 (0.21)

Table 4: Bookmarked and read documents by topic

Table 4 shows how these behaviours and the performance varied by topic, again indicating that the pupils found topic 353 to be particularly difficult, only managing to bookmark a relevant

document 11% of the time. Equally notable, however, is relatively how well they performed on topic 367 where 63% of the documents they chose to bookmark were relevant.

Measure	Q1	Q4
# bookmarked	1.03	1.0
# read	1.58	1.66
# relevant bookmarked (ratio)	0.36 (0.35)	0.38 (0.38)
# relevant read (ratio)	0.4 (0.26)	0.33 (0.2)

Table 5: Bookmarked and read documents by quartile

Overall the Q4 pupils read more documents but were also more likely to read a non-relevant document, however, while they tended to bookmark slightly less often, they were more accurate in doing so. 38% of the bookmarks by the 4th quartile pupils were relevant, while only 35% of those bookmarked by the 1st quartile pupils were relevant. Although not entirely conclusive, this suggests that the 4th quartile were able to use the information gained from their extra reading of documents to more accurately determine which were relevant.

Task time

On average, participants spent 472.5 seconds (7.88 minutes; median = 535 seconds / 8.92 minutes) completing each task. This varied between a minimum of 80 seconds and a maximum of 884 seconds (14.73 minutes). Duration by topic was broadly the same, although participants did tend to spend slightly longer on topic 408 (*tropical storms*; median = 573s) and slightly less time on topic 347 (*wildlife extinction*; median = 493). Unsurprisingly, participants generally spent less time on their second topic (461 seconds) than on their first (494.3 seconds). There was very little difference in terms of topic duration between 1st and 4th SIMD quartile pupils (median = 562 and 561 seconds respectively).

Query formulation

Research has shown that search experts (who achieve excellent performance from search systems) typically formulate queries that have consistent characteristics. Their submitted queries are quite long, being composed of multiple keyword terms, use domain-specific vocabulary, are correctly spelt and tend not to copy text verbatim from the information need/TREC topic (Harvey et al, 2015)). Furthermore, they often submit multiple query reformulations, learning from the results of their earlier queries to improve their success/performance (Aula et al, 2005; White and Morris, 2007). In this section we consider the characteristics of the pupils' queries to understand how (dis)similarly they are behaving to expert users.

Query complexity

Participants submitted an average of 5.4 queries per topic (median=4), although one particularly keen pupil submitted a very large number of 25 queries for a single topic. This is somewhat less than reported in other studies by Torres et al. (8.76 queries; Torres et al., 2010) and Bilal's study of children's use of Yahoo!igans! (6.7 queries; Bilal, 2002). Users have been found to become discouraged when a search engine does not immediately return good results (Pickard, 2004). Since our search system excluded features such as spelling autocorrection and suggested queries by design, this could go some way to explaining the lower number of queries we observed.

The average query length was 3.5 terms (median=3), which is approximately equal to other studies, which report an average between 2.84 and 3.23 (Torres et al., 2010; Duarte Torres and Weber, 2011). The average character length was 24.9 (median=22) characters, which is slightly

longer than those of the 13 to 15 year olds' queries in the Duarte Torres and Weber paper (2011), which reported 17.67 characters for the non-navigational task. Linear modelling confirms that longer, more specific queries result in better performance – both the number of query terms and the character length of queries are strong predictors of result quality in terms of both number of hits and average precision. An increase in query length of a single term yields an expected 0.012 improvement in average precision ($p < 0.01$) and the addition of 0.42 extra hits ($p < 0.01$).

Topics 347 and 408 had longer queries, with more characters and more time spent formulating those queries, while queries for topic 367 were noticeably shorter and less time was spent on this topic. Analysis of linear models indicates that pre-task interest in the topic has a significant impact on the term (coef.=0.15, $p = 0.04$) and character length (coef.=1.87, $p < 0.01$) of the queries submitted for that topic. This suggests that the participants spent more time thinking about and formulating the terms which made up their queries for those topics they were interested in.

Topic	# terms	# chars	# queries	Duration (s)
347 – Wildlife extinction	4	24	3	54.5
353 – Antarctica exploration	3	22	4	38.5
367 – Piracy	3	19	4	33
408 – Tropical storms	4	26	4	44

Table 6: Querying statistics by topic (median values)

Although not significant, participants in the 4th SIMD quartile tended to submit longer queries than those in the 1st quartile (mean of 3.7 terms against 3.2 terms) and spent longer querying (mean duration of 62.3s against 57.7s).

Mistakes, use of advanced search and off-topic queries

The instance of mistakes (mostly misspellings) was high at 118 or 18.89% of all of the 625 queries, which is a larger proportion than has been found previously - Bilal observed 2% of mistakes (Bilal, 2002). A considerable proportion of these were due to the participants incorrectly spelling 'Antarctica', although mistakes were observed for all four topics.

Unsurprisingly, mistakes have a significant correlation with query performance ($p < 0.01$), where an increase of 1 mistake has a -0.031 effect on average precision.

Overall, there are very few instances of advanced operators being used, with a total of 14 instances or 2.24% of all queries containing some form of advanced operator. There 25 off-topic queries, which equates to 4% of all queries. Foss et al. (2013) discuss adolescents as being 'more aware of social expectations placed on them when participating in a research study, and are more likely to answer questions directly'. In the context of this study, the participants' search queries were at times off-topic ('aidan denholm', 'beach volleyball'), or demonstrated some frustration with the search system ('please work') and, at times, crossed the line into vulgarity ('f**k this'). Many of these queries were submitted for topic 367 ('piracy'), perhaps again reflecting the pupils' general disinterest in this particular topic.

Lack of assistive functionality is in line with that of Bilal (2002), and although its adoption by many youth-oriented search engines, such as Yahoo!igans!, is advocated by Bilal, it is worth noting that not all search facilities outside of web search incorporate such features. This raises the question about whether this should be adopted wholly or is there a means to support children, and adults (who struggle just as much with the lack of assistance) in some other way without such functionality?

By quartile

Measure	Q1	Q4
Mistakes	0.17	0.28
Advanced operators	0.01	0
Off-topic	0.1	0.03
Query length (terms)	3.23	3.66
Query length (characters)	23.2	26

Table 7: Querying characteristics and mistakes by quartile

Perhaps surprisingly the fourth quartile pupils submitted a significantly larger number of queries with mistakes than those in the first quartile, although this may be explained by them attempting to construct longer, more elaborate queries. First quartile participants submitted over 3 times as many off-topic queries, perhaps reflecting a greater tendency to become frustrated by the lack of querying support offered by the system. The fourth quartile pupils submitted longer queries as counted by both number of terms and character length, suggesting that they provided better-specified, more detailed queries.

*Post task**Questionnaire*

Perception of task difficulty increased post task to 2.92 from 2.88 pre-task. Participants found it more difficult to find relevant information across all tasks, with the largest increases between pre and post responses for topic 347 and 353, (2.625 to 3.313 and 2.833 to 3.500 respectively). As expected, they found identifying keywords most difficult for task 353 (mean 3.583) and least difficult for 408 (mean 2.613).

Despite this mixed perception of difficulty, the users were able to understand the information they read with 408 highest at 4.000 and 353 lowest at 3.375 (overall mean 3.77) and trusted the information found (highest 3.250 for topic 347 and lowest 2.968 for topic 408, overall mean 3.14) and were reasonably satisfied across all tasks that they had found the information required to complete the task (highest 347, 3.438 and lowest 353, 3.042, overall 3.28).

Linear modelling revealed that self-reported task knowledge, post-task difficulty and self-reported ability to understand the content of retrieved documents were not significant predictors of actual retrieval performance, which was also identified by Bilal (2002). That said, post-task satisfaction with the information found had a significant positive effect on performance, where an increase of 1 increased precision by 0.012 ($p=0.030$).

In previous studies young people were found to overestimate their ability (Pickard, 2004) or be unable to identify when they have performed badly (Schacter et al. 1998). Despite being able to identify when a task was more difficult, trust and understanding of the information were generally quite high and overall satisfaction in what had been found (and bookmarked) was also high, indicating that the pupils in this study also tended to overestimate their abilities and failed to ascertain when they had performed poorly.

Post task responses

Post interview responses show the users' understanding of the concepts of search, and the justification for decisions made. However, despite this understanding it further emphasises the shortcomings in their information literacy abilities and awareness of said shortcomings. For example, when discussing the reason for selecting a particular document, one user identified it was because '*it was most relevant to the question*'. Although this particular user's performance

for this topic was 0.132 which in itself is quite low, it was almost double that of the average for that topic, which was 0.069. Other comments regarding document selection follow in a similar fashion: *'probably contained relevant information'*, *'stood out as having relevant information'*, *'[was a] detailed article'*, *'summed up effects of piracy'*, *'seemed interesting'*, *'timely document that seemed relevant'* and *'stood out as having relevant information'*. While these comments demonstrate a simplistic understanding of topic and the document content, they do not suggest that the pupils either possess strong information literacy skills or that they do not take the time to employ them. If the users can differentiate information based on such distinctions then one would expect their performance to be better, so it is clear there is a disconnect, but also that their post-task perception that they are able to identify and select the most relevant sources of information is inaccurate and overly optimistic.

When discussing query submission and query formulation, users adopted mixed strategies: *'based upon words from the question'*, *'breaking the question down into keywords'*, *'related to the initial search'*, while others *'tried new things (nothing worked)'*, *'tried new ideas'* and *'tried to find as many different things as possible'*. Search strategies employed range from keyword search, paraphrasing the initial question and reformulation. What is interesting is that in the instance of *'new things'*, which didn't work, the researcher noted that this user was reformulating, appending to the initial term without correcting the reason for the poor retrieval performance – that the first term was spelled incorrectly.

Conclusion

The results of this study demonstrate a lack of good application of information literacy and search skills by the participants. Their typical performance, both in terms of query quality and ability to identify relevant information, is quite poor. Low performance has been attributed to lack of interest or knowledge of a topic (Smith and Hepsworth, 2007), however, in this case interest was reasonably high, and, although topic knowledge was reported at a low level, this was not found to have a significant negative impact on performance. What is particularly striking is that participants often selected (bookmarked) documents that were not relevant for their assigned topics and struggled to modify their queries to improve performance. Despite this poor performance, the pupils generally felt they had performed well and did not consider the tasks to be especially difficult, suggesting a lack of awareness of their own performance and skills development.

The results also highlight a reliance of the pupils on search assistance functionality, such as spellchecking and suggested queries, which are often present in major search engines but are by no means implemented by all search systems.

There were some marked differences between the SIMD quartiles, with fourth quartile students reading more documents, submitting larger more complex queries and subsequently more mistakes, and performing better when the topic is well known and interesting. However, overall performance and time spent on task were approximately equal. It must, therefore, be acknowledged that despite potential advantages living in a less deprived area may provide, this does not appear to equate to success in learning.

Acknowledgements

This research was completed with the support of a bursary from the Information Literacy Group (ILG) of the Chartered Institute of Library and Information Professionals (CILIP). The studies were conceived and conducted in partnership with colleagues from the Scottish Information Literacy Community of Practice and Edinburgh City Libraries and Schools. The authors would like to thank Katie Swann, Cleo Jones, Ian McCracken and David Hastings for their help with the project and the pupils and headmaster, Tom Rae, at Craigmount High School in Edinburgh.

References

- Aula, A., Jhaveri, N. & Kaki, M. (2005). Information search and re-access strategies of experienced web users, WWW '05, ACM, pp. 583–592.
- Bilal, D. (2000). Children's use of the Yahoo!igans! Web search engine: I. Cognitive, physical, and affective behaviors on fact- based search tasks. *Journal of the Association for Information Science and Technology*, 51(7), pp.646-665.
- Bilal, D. (2002). Children's use of the Yahoo!igans! Web search engine. III. Cognitive and physical behaviors on fully self- generated search tasks. *Journal of the Association for Information Science and Technology*, 53(13), pp.1170-1183.
- Bilal, D. (2013). Comparing google's readability of search results to the flesch readability formulae: a preliminary analysis on children's search queries. *Proceedings of the Association for Information Science and Technology*, 50(1), pp.1-9.
- Borlund, P., (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research. An International Electronic Journal*, 8(3).
- Coiro, J. and Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet, *Reading research quarterly* 42 (2), pp. 214-257.
- Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M. L., Fails, J., & Hutchinson, H. (2009, June). How children search the internet with keyword interfaces. In *Proceedings of the 8th International conference on interaction design and children*. pp. 89-96. ACM.
- Duarte Torres, S., Hiemstra, D., & Serdyukov, P. (2010, July). Query log analysis in the context of information retrieval for children. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. pp. 847-848. ACM.
- Duarte Torres, S. & Weber, I., (2011). October. What and how children search on the web. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. pp. 393-402. ACM.
- Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information Problem Solving: The Big Six Skills Approach to Library & Information Skills Instruction*. Ablex Publishing Corporation, 355 Chestnut St., Norwood, NJ 07648.
- Foss, E., Druin, A., Yip, J., Ford, W., Golub, E. & Hutchinson, H., (2013). Adolescent search roles. *Journal of the Association for Information Science and Technology*, 64(1), pp.173-189.
- Hahnel, C., Goldhammer, F., Kröhne, U. & Naumann, J., (2018). The role of reading skills in the evaluation of online information gathered from search engine environments. *Computers in Human Behavior*, 78, pp.223-234.
- Harvey, M., Hauff, C., Elswiler, D., (2015). Learning by Example: training users through high-quality query suggestions. *Proceedings of the 38th Annual ACM SIGIR Conference*. Santiago, Chile.
- Harrop, C. & Seddon, L. (2012). Competent and confident, CILIP Update. pp. 44-45.
- Helbig, N., Gil-García, J. R., & Ferro, E. (2009). Understanding the complexity of electronic government: Implications from the digital divide literature. *Government Information Quarterly*, 26(1), pp.89-97.
- Pickard, A. J., Shenton, A. K. & Furness, K. (25th- 28th June 2013) Educating young people in the art of distrust: Meta-evaluation and the construction of personal, agile models of web information literacy. 'i3: Information, impacts and interactions.' Robert Gordon University June 2013.

- Pickard, A. J., Shenton, A. K., & Johnson, A. (2014). Young people and the evaluation of information on the World Wide Web: Principles, practice and beliefs. *Journal of Librarianship and Information Science*, 46(1), pp.3-20.
- Rowlands, I. & Nicholas, D. (2008). Information behaviour of the researcher of the future: A cyber briefing paper. Retrieved from https://www.webarchive.org.uk/wayback/archive/20140614113419/http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf. (Archived by WebCite® at <http://www.webcitation.org/6NzsAsSou>).
- Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H.R., Dobrowolski, T. & Tenopir, C., (2008, July). The Google generation: the information behaviour of the researcher of the future. In *Aslib proceedings*. 60 (4), pp. 290-310. Emerald Group Publishing Limited.
- Schacter, J., Chung, G.K. & Dorr, A., (1998). Children's Internet searching on complex problems: performance and process analyses. *Journal of the Association for Information Science and Technology*, 49(9), pp.840-849.
- Scottish Government (2016) Introducing the Scottish Index of Multiple Deprivation. Retrieved from <http://www.gov.scot/Resource/0050/00504809.pdf>. (Archived by WebCite® at <http://www.webcitation.org/6xXwg2xRa>).
- Shenton, A. K. & Pickard, A. J. (2014). Evaluating online information sources. Minibook 42. Leicester: The United Kingdom Literary Association.
- Smith, M., & Hepworth, M. (2007). An investigation of factors that may demotivate secondary school students undertaking project work: Implications for learning information literacy. *Journal of Librarianship and Information Science*, 39(1), pp.3-15.
- Sturges, P., & Gastinger, A. (2010). Information literacy as a human right. *Libri*, 60(3), pp.195-202.
- White, R. W., & Morris, D. (2007, July). Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* pp. 255-262. ACM.
- Yu, L. (2006). Understanding information inequality: Making sense of the literature of the information and digital divides. *Journal of Librarianship and Information Science*, 38(4), pp.229-252.