



# Student digital information-seeking behaviour in context

David Nicholas and Paul Huntington

*Centre for Information Behaviour and the Evaluation of Research (CIBER),  
School of Library, Archive and Information Studies,  
University College London, London, UK*

Hamid R. Jamali

*Department of Educational Technology, Tarbiat Moallem University,  
Tehran, Iran, and*

Ian Rowlands and Maggie Fieldhouse

*Centre for Information Behaviour and the Evaluation of Research (CIBER),  
School of Library, Archive and Information Studies,  
University College London, London, UK*

## Abstract

**Purpose** – This study provides evidence on the actual information-seeking behaviour of students in a digital scholarly environment, not what they thought they did. It also compares student information-seeking behaviour with that of other academic communities, and, in some cases, for practitioners.

**Design/methodology/approach** – Data were gathered as part of CIBER's ongoing Virtual Scholar programme. In particular log data from two digital journals libraries, Blackwell Synergy and OhioLINK, and one e-book collection (Oxford Scholarship Online) are utilized.

**Findings** – The study showed a distinctive form of information-seeking behaviour associated with students and differences between them and other members of the academic community. For example, students constituted the biggest users in terms of sessions and pages viewed, and they were more likely to undertake longer online sessions. Undergraduates and postgraduates were the most likely users of library links to access scholarly databases, suggesting an important "hot link" role for libraries.

**Originality/value** – Few studies have focused on the actual (rather than perceived) information-seeking behaviour of students. The study fills that gap.

**Keywords** Students, Information research, Information retrieval, Electronic journals, Transaction analysis

**Paper type** Research paper

## Introduction

The information seeking behaviour of students has been the subject of much debate in recent years as the mass availability of information on the web has led to widespread concerns about plagiarism (*BBC News*, 2006), the unthinking, unevaluated, over-usage of web resources by students (Graham and Metaxas, 2003) and, more generally, as researchers and practitioners wonder whether a fundamental shift in searching for and researching content has occurred amongst young people. This is a critical issue for higher education institutions. Brabazon (2007) describes in her book "The University of



---

Google” how education systems somehow confuse access to digital information with developing informed citizens. Her book is an attempt to show how education systems in the information age should enable students to take a journey through knowledge, rather than being consumers in the shopping centre of cheap ideas.

However, the literature tends to be long on speculation and light on detail, over dependent on self-report methods and parochial (relating to individual journals, rather than student communities). It also often lacks a context and comparison and without this it would not be possible to establish whether student behaviour is any different to that of any other scholarly group. In order to address these weaknesses, log data obtained from the four year long Virtual Scholar evidence-based research programme ([www.ucl.ac.uk/slais/research/ciber/virtualscholar/](http://www.ucl.ac.uk/slais/research/ciber/virtualscholar/)) has been evaluated from the perspective of the academic status of the scholar, which not only provides a comprehensive evaluation of the information seeking behaviour of the student, but also compares it with that of other members of the academic community – professors, teacher and researchers, and in some cases with that of practitioners. It is thought that this is the first time that log data has been used to provide a contextual understanding of students’ information seeking behaviour. This is because traditional log analyses cannot distinguish between the use of different academic user groups using the same platform, but the method employed here (deep log analysis) can.

This paper emanates from a British Library/JISC funded project, “Study on the Information Behaviour of the Researcher of the Future” or “Google Generation” (CIBER, 2007), which sought among other things to discover how young people conducted their online research and whether this was different in anyway from that of older researchers and scholars.

### **Aims/objectives**

The aim of this paper is to:

- provide robust evidence on the actual (rather than reported) information seeking behaviour of students (undergraduates and postgraduates) in a digital scholarly environment by mining and consolidating the log data (more than three million transactions) that have been generated by the Virtual Scholar research programme over the past four years;
- compare student information seeking behaviour with that of other academic communities, and in some cases, practitioners; and
- wherever appropriate, relate and contrast log based information seeking data with the findings of other researchers who have employed self-report (surveys and interviews) and traditional forms of transactional log analysis.

### **Virtual scholar log studies sourced**

The following Virtual Scholar studies yielded data on students’ use of scholarly databases, covering e-journals and e-books, and constitute the database on which this paper is built:

- An evaluation of the use of Synergy, a digital journal database produced by Blackwell and covering more than 700 journals. For this investigation a little over 500,000 user transactions to the Blackwell site were analysed. Transactions relate to a single 24 hour period for the 17 September 2003.

- An evaluation of OhioLINK, the database of a consortium in Ohio offering more than 6,000 journals to its users. The raw logs were obtained from OhioLINK for four universities regarding on-campus use for the fifteen month period January 2005 to April 2006. Over 2,250,000 transactions were analysed.
- An evaluation of Oxford Scholarship Online (OSO), a collection of over 1,200 e-monographs conducted as part of the SuperBook study at University College London (More information projects can be found at [www.ucl.ac.uk/slais/research/ciber](http://www.ucl.ac.uk/slais/research/ciber)). The study was undertaken between October and December 2007, and 4,240 transactions were analysed.

While these studies were conducted at different times, on different databases and students were identified in different ways, they all had a similar aim, which was to establish and evaluate the information seeking behaviour of the virtual scholar and were unified by the fact they used exactly the same methodology a variant of transactional log analysis, deep log analysis.

### **Literature review**

Highlighting the importance of this study, it has only been possible to identify one log study that has investigated usage by students and that was conducted more than ten years ago, when the internet was still very much in its infancy. This was TULIP (The University Licensing Project), a collaborative project carried out between Elsevier Science and nine universities in the USA during the period 1991-1994. Although the study's primary focus was e-journal delivery and usability issues, log file analysis yielded data showing that graduate students viewed more abstracts and searched electronic journals more actively and with a broader focus than faculty. While undergraduate usage was not examined in depth, log data also indicated a significant degree of activity by this group. (Borghuis, 1996)

Therefore the literature review that follows is largely concerned with the findings of studies based on surveys and interviews which concern themselves with student information seeking in a digital environment. The literature shows that undergraduate students opt for the easiest and most convenient method of information seeking (Valentine, 1993), and appreciate the time saving characteristics of electronic resources (Dalglish and Hall, 2000). Students are said to rely heavily on simple search engines, such as Google to find what they want. (Dalglish and Hall, 2000; Becker, 2003; Drabenstott, 2003). Fast and Campbell (2004) undertook an exploratory study of how university students perceive and interact with web search engines compared to web-based OPACs. A qualitative study was conducted involving just sixteen students, eight of whom were first-year undergraduates and eight of whom were graduate students in library and information science. The participants performed searches on Google and on a university OPAC. The interviews and "think-afters" revealed that while students were aware of the problems inherent in web searching and of the many ways in which OPACs are organized, they generally preferred web searching. The coding of the data suggests that the reason for this preference lies in psychological factors associated with the comparative ease with which search engines can be used, and system and interface factors which made searching the web much easier and less confusing.

Bilal (1998, 2000) explored seventh grade students' use of Yahoo!igans!, a web search engine designed specifically for young people. The young web users tended to

---

examine briefly the first few hits on the initial results pages before performing new searches, rather than examining every hit in detail. This is termed a “satisficing” (a combination of words “satisfying + sufficing”) form of behaviour that enables users to deal with prohibitively large amounts of information. Satisficing acts as a “stop-rule” (Simon, 1979, p. 4), once an acceptable alternative is found, the decision maker concludes the decision process. Nonetheless, satisficing does not limit the decision maker to one deciding factor, nor does it lock the decision maker into searching for an unrealistically superior option or a needlessly inferior option. Bilal’s (2000) study showed that the study participants preferred keyword searching to browsing, because this reduced more rapidly the pool of sites from which they must make selection decisions. Another study on students which relates to the concept of satisficing is that conducted by Prabha *et al.* (2007). They showed that undergraduate and graduate students tend to stop looking for information when they find the required number of sources for an assignment. These findings add weight to those of Barrett (2005) who found that undergraduate students sought to find enough information to fulfill course requirements.

Schacter *et al.* (1998) found fifth- and sixth-grade students had difficulty finding desired information on the web. They also found that participants exhibited a strong preference for browsing over analytical (planned or pre-meditatively structured) search techniques. Fidel *et al.* (1999) studied high school students’ web-searching patterns using observation and think-aloud protocol analysis. The participants were quick to abandon seemingly unsuccessful searches, returning to known landmarks to begin new searches. A recent extensive review of the relevant literature by Rowley and Urquhart (2007) indicates that there are gaps in the evidence concerning the browsing and selection strategies of undergraduate students and the interaction of some of the mediating influences on information behaviour.

Important factors affecting information behaviour of students seem to be the type of learning task (Kerins *et al.*, 2004), teaching and learning styles (Eskola, 2005), the motivation to learn and personality (Heinström, 2005). Urquhart and Rowley’s (2007) study deals with factors that affect student information behaviour. They list two sets of factors that influence students’ information seeking behaviour:

- (1) *Macro factors*: information resource design, information and learning technology infrastructure, access, organizational leadership and culture, and policies and funding.
- (2) *Micro factors*: information literacy, search strategy, academics’ role in changing information behaviour, discipline and curriculum, pedagogy, support and training.

A number of writers have demonstrated diversity in regard to student information seeking, especially in regard to discipline (Entwistle, 2003; Whitmire, 2002). Whitmire used the Biglan model of disciplinary differences (dimensions of hard, soft, pure, applied, non-life, and life), and found some significant differences between disciplines in a large questionnaire-based study of undergraduate students (5,175 respondents). She found that undergraduates in the soft disciplines (humanities, business, social sciences, and education) compared to undergraduates in hard disciplines (physical sciences and engineering) engaged in more information-seeking activities, with the exception of using the library as a place to read or study. The differences in the

information-seeking behaviour patterns were statistically significant for seven of the ten activities: using the on-line catalogue, asking the librarian for help, reading in the reserve/reference section, using indexes to journal articles, developing a bibliography, checking citations in documents read, and checking out books. When comparing pure disciplines (physical sciences, humanities, and social sciences) with applied disciplines (engineering, business, and education), it was found that undergraduates in the pure disciplines engaged in more information-seeking activities when compared to students majoring in the applied disciplines.

The JUSTEIS project, a project which sought to examine the uptake and use of electronic information services in higher education in the UK, also uncovered disciplinary differences (Urquart *et al.*, 2003). It showed that:

- the clinical medicine and biological sciences community used and valued electronic information sources most; and
- humanities and arts undergraduates ranked search engines as their most frequently used electronic resources.

More generally, the study showed the increasing popularity of electronic journal services, the acceptance of the search engine model for information retrieval and the important role academic staff play in the promotion of electronic information services for student learning.

Whitmire (2003) used Kuhlthau's Information Search Process together with four models of epistemological development, and identified different patterns of behaviour amongst undergraduates according to the level of epistemological belief, lending further support to the intertwining of learning, discipline, and approaches to knowing what one knows. Liu and Yang's (2004) questionnaire survey study of distance-learning students also found significant relation between the selection of information resources and their subject discipline. The study showed that the field of study was a good predictor for the respondents' use of the libraries databases/e-journals; and more importantly, motivation was significantly related to the respondents' field of study. More of the highly motivated respondents chose the libraries as their primary information source than did those choosing the internet or the other sources. Less-motivated respondents tended to take the internet and other resources as their primary sources. In Urquhart and Rowley's (2007) longitudinal study, the aggregated questionnaire data showed that there were disciplinary differences in the use of electronic journals, with significantly greater use among students in the clinical disciplines, and pure and applied sciences, and less use among Humanities and arts, and pure and applied social sciences

Urquhart and Rowley (2007, p. 1192) on the basis of an extensive analysis of the literature, characterised student behaviour in the following terms: "first-year undergraduates indicated that the route they chose to finding information was governed by time factors, convenience of format, and an unwillingness to try the unfamiliar unless this was an explicit expectation", which may seem at odds with what young people are assumed to be doing more generally in the virtual environment; trail-blazing social networking and the like. However they go on to say that "Undergraduates who had progressed beyond the first year were more likely to mention some other quality criteria such as currency of information, the reliability of the source, and the authority of the source, but time saving was important for them as well."

The limitation of many of the studies reviewed here is that they draw very broad portraits of student information seeking. While this is helpful in understanding certain aspects of information seeking behaviour of students, such as their motivations, another type of very important information that policy makers and system designers need is highly specific data, data on the number of documents retrieved, amount of time spent online and types of material viewed. This is the area in which the current paper seeks to contribute. Also few studies examine student information seeking in terms of the information seeking behaviour of other scholarly groups, something both librarians and publishers are very interested in.

### Research methods

Digital information platforms have a facility by which logs are generated that provides an automatic and real-time record of use by anyone who uses them. They represent the digital information footprints of the users and by analysing the transactional logs information seeking behaviour can be mapped. When these footprints are enhanced with user data, for instance, by linking to user databases and online questionnaires- we call this deep log analysis (DLA); they then can say something about the kinds of people that use the services and how their behaviour differs, in the case of the paper differences amongst types of scholars.

While logs are attractive, not least in that in that they provide data on what very large numbers of people did and not what they say they did (as is the case with survey studies), they do have their weaknesses. Thus logs do not record all user transactions (caching is a problem), user session information can become muddled (proxy-servers are the problem) and because of dynamic IP addresses users are difficult to define. For more information on problems regarding log analysis see Jamali *et al.* (2005). Hence, logs can only offer an indication of likely underlying behavioural patterns that need to be followed up by follow-up surveys and more qualitative methods. They raise the questions that need to be answered. For more details of the methodology see Nicholas *et al.* (2000, 2006)

The logs referred to in this paper were standard logs and detailed the internet protocol (IP) address, date, time and the page viewed by a user. The following OhioLINK log is a typical example of the kind of logs analysed (see Figure 1).

The first field (129.22.7.22) details the user's internet protocol (IP) number. Users are not required to enter a user name or password to enter OhioLINK. OhioLINK authenticates entry based on the user's specified fixed IP addresses or range of IP addresses. The second field [31/May/2004:00:06:39 -0400] details the date and time of the transaction and how many hours off Greenwich meantime it is. The third field details the item delivered in response to a request from the user: "GET /cgi-bin/sciserv.pl?collection = journals&journal = 02663538&issue = v61i0006 &article = 889\_ tdpogerdp&form = pdf&file = file.pdf. The number 02663538 is the ISSN number of the journal, although the hyphen after the fourth digit is missing. The number v61i0006, in the logs, specifies the volume and issue number – in this case the

```
129.22.7.22 - - [31/May/2004:00:06:39 -0400] "GET /cgi-bin/sciserv.pl?collection=journals
&journal=02663538&issue=v61i0006&article=889_tdpogerdp&form=pdf&file=file.pdf
HTTP/1.1" 302 5 "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; .NET CLR .1.4322)" "-"
```

Figure 1.



number relates to volume 61 issue 6. The alpha-numeric sequence 889\_tdpogerdp identifies the article. Lastly “file.pdf” denotes that the file was supplied in a PDF format, in fact almost all article documents were supplied as PDF. The fourth field details the status of the download. The delivery status field records whether the page was delivered without any problems. The fifth field records the browser details and is the record of the browser that the client is using to attach to the internet.

Logs normally do not provide more than a trace of the user’s identity as evidenced by an IP number but using deep log techniques it was possible to relate usage to academic status by various means. In the case of the Synergy study it was possible to link usage logs to a subscriber identifier which enabled demographic data to be related to usage, and academic status was one of the demographics furnished. In the case of the two other studies (OhioLINK and OSO) students were identified by the sub-network they used. This latter method is somewhat more problematic because it is not possible to be exactly sure from the name of the network that only students (or staff) used these networks, although the fact that searchers were searching, for instance, from the halls of residence or the library meant that it was highly likely that most users were students.

Logs were processed using the Statistical Package for the Social Sciences (SPSS).

### Results

Student information seeking was evaluated in regard to three databases – Synergy, OhioLINK, and Oxford Scholarship Online and Table I provides a summary of the ways in which information seeking was characterised in the case of each database.

#### Synergy study

The analyses presented here are based on data that were collected for all those users who could be related to the database of subscribers Blackwell’s maintained at the time (autumn 2003). The occupational status of the user was derived from the user response form and this resulted in the following categories: professors and teachers (24 per cent of users identified), researchers (23 per cent), postgraduates (19 per cent), undergraduates (12 per cent) and professionals/practitioners (22 per cent). In most academic institutions students would constitute the majority academic community and this proved also to be the case in usage terms (see below).

Characteristic	Synergy	OhioLINK	Oxford Scholarship
Number of page views	✓	✓	✓
Time online		✓	✓
Site penetration	✓	✓	✓
Type of page viewed	✓	✓	
Referrer link used	✓		
Searching/navigating to content	✓	✓	
Subject of document viewed		✓	
Age of document viewed	✓	✓	
Number of journals viewed	✓	✓	
Added functions used	✓		

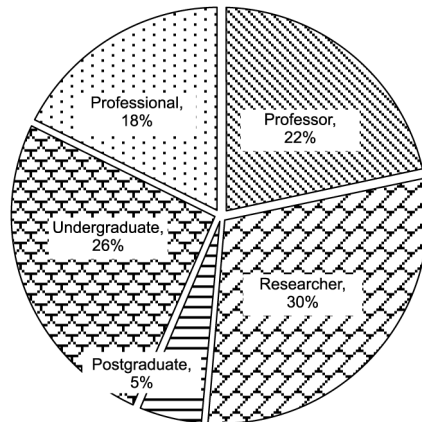
**Table I.**  
Characteristics used to  
profile information-  
seeking behaviour

*Levels of activity (number of sessions conducted and number of pages viewed)*

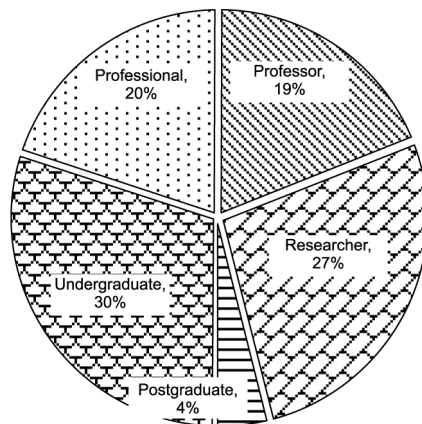
Postgraduates accounted for 5 per cent of the online sessions conducted and 4 per cent of the page views and undergraduates 26 per cent of sessions and 30 per cent of views. Together we have a student figure of 31 per cent (sessions) and 34 per cent (page views), which meant that students were the biggest user group. By way of contrast, professors and teachers accounted for 22 per cent of sessions conducted and 19 per cent of all views. Overall, usage declined with academic status, which is unsurprising, given the smaller numbers of people involved and the greater networking opportunities available to more senior staff. Figures 2 and 3 relate.

*Site penetration (number of page views in a session)*

In terms of heavy use (classified as viewing 21 or more pages in a session), the likelihood of being a heavy user increased with academic status. Thus, for undergraduates, just 7 per cent of the sessions conducted were heavy use sessions; this was the case with 10 per cent of postgraduate sessions, 12 per cent of researcher sessions and 11 per cent of professor/teacher sessions. A similar pattern was found for



**Figure 2.**  
Percentage distribution of  
sessions by occupational  
status



**Figure 3.**  
Percentage distribution of  
page views by  
occupational status



medium use sessions, those sessions consisting of 11 to 20 page views: the percentages were 13 per cent for undergraduates, 12 per cent for postgraduates, 19 per cent for researchers and 18 per cent for professor/teacher sessions. The information seeking profile of undergraduates as a group is that, partly as a function of their numbers, they conduct many sessions and view many pages but do not penetrate web sites very deeply during their visits. This is what has led to them being called “bouncers”, they bounce in and then bounce out again (Nicholas *et al.*, 2007) (see Figure 4).

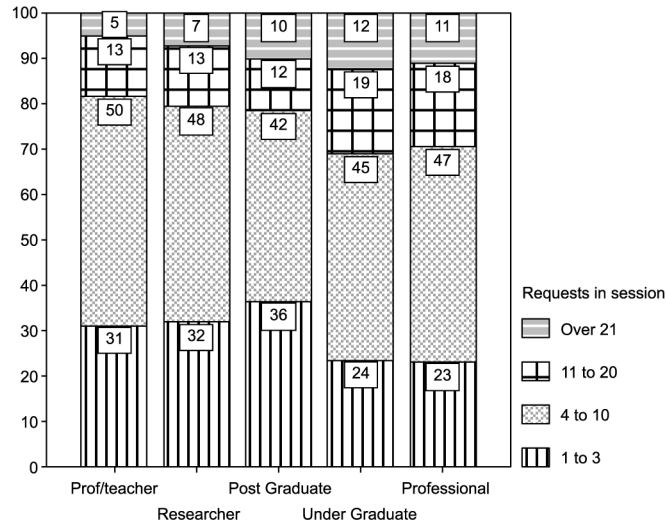
*Type of page viewed*

In terms of the type of page viewed, surprisingly perhaps, undergraduates proved to be the biggest viewers of abstracts – 24 per cent of page views were to abstracts. The use of PDFs increased as users moved up the academic scale, that is use of PDFs increased as the user moved from undergraduate (19 per cent) to postgraduate (25 per cent) to researcher (26 per cent) and then to professor/teacher (30 per cent). Perhaps undergraduates were much more interested in cutting and pasting, something much easier to do in HTML format? (see Figure 5).

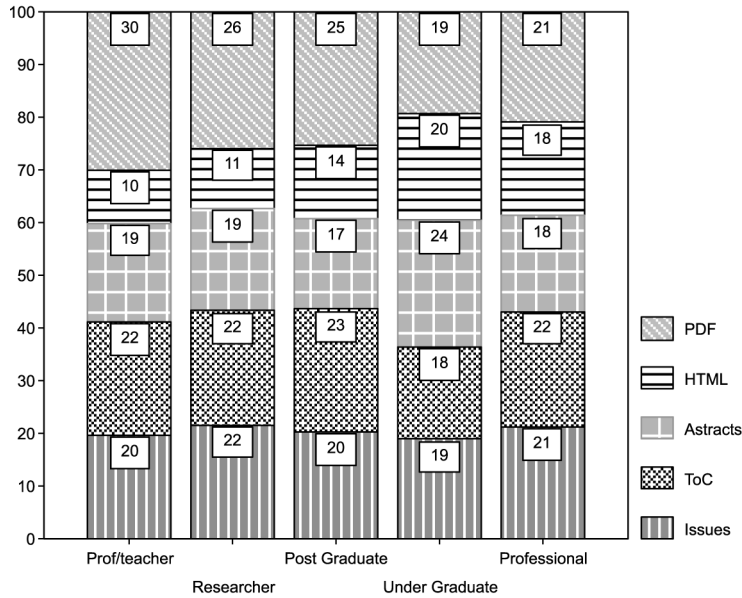
Table II gives the page view time for various types of pages. Undergraduates took much longer to view almost every type of page with the exception of pop-ups, and, significantly, they were the fastest viewers on this count. What particularly stood out is how much longer they took to view a PDF, they recorded an average time of 71 seconds, about 70 per cent longer than expected (42 seconds), which suggests that they were having download problems or actually preferred to read online this way (and the OhioLINK study reported later provides support for the latter hypothesis).

*Referrer link used*

Figure 6 gives the distribution of type of referrer link used (the site from which people arrived) by the academic/occupational status of the user. It shows that undergraduates (7 per cent) and professionals (5 per cent) were the predominant users of journal web



**Figure 4.** Number of page views in a session by occupational status (%)



**Figure 5.** Type of page viewed by organisational status (%)

	Teacher/Professor	Researcher	Postgraduate	Undergraduate	All users
ToC	24.8	20.5	23.9	32.6	24.0
Abstract	27.8	26.1	20.9	28.1	25.9
Full text HTML	20.4	31.3	26.0	21.8	31.7
Full text PDF	20.8	32.6	31.1	71.2	42.3
Pop-ups	55.6	70.6	67.8	55.1	43.8
All	23.0	27.1	24.7	32.3	28.4

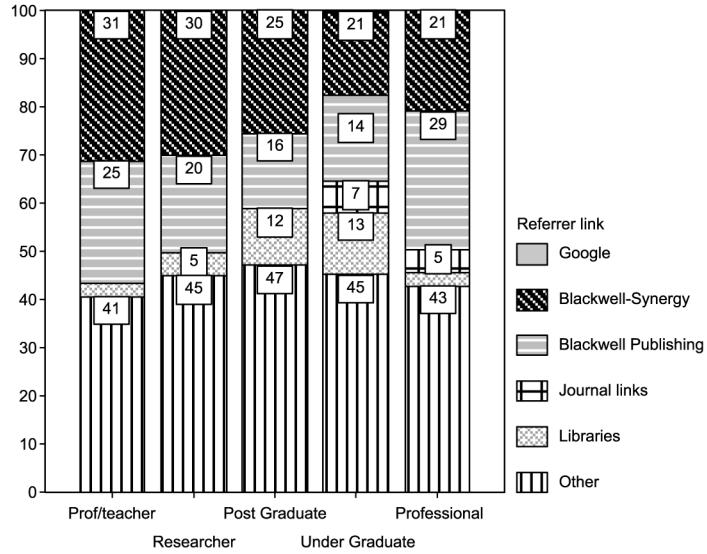
**Table II.** Time taken to view a page, in seconds

**Note:** All estimates are Hubers-M estimate of the mean

links. Undergraduates and postgraduates were the most likely users of library links; 13 per cent and 12 per cent of members of these user groups did so compared to about 5 per cent for other groups. Professors and teachers were the most likely to access Synergy directly (31 per cent) and via the Blackwell Publishing main site (25 per cent). During this period search engines were debarred from indexing the site so do not feature in the analysis, although some users did find the site via Google there were very few users who found the site this way and they represented less than half a percent of all use.

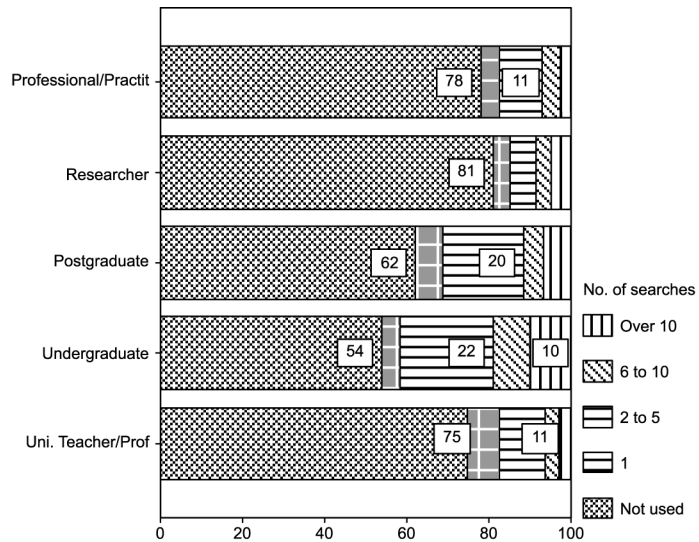
### Search preference

Most Synergy sessions did not feature the use of the internal search facility but of those that did undergraduates, as might have been expected, were the most likely to use it (46 per cent did so), while researchers (19 per cent), professional/practitioners (22 per cent) and teachers/professors (25 per cent) were much less likely to do so.



**Figure 6.**  
Type of referrer link used  
by occupational status

A further analysis identified how many individual searches were conducted in a session where the on-site search facility had been employed. Undergraduates undertook the greatest number of searches, 10 per cent of all sessions saw more than ten searches being conducted. What is not clear is whether this constituted effective searching or not (see Figure 7).



**Figure 7.**  
Number of searches  
conducted in a session by  
occupational status

*Age of documents viewed*

Undergraduates were more likely to view older articles and postgraduates more recent ones. In the case of undergraduates the explanation, perhaps, could be that there is a lag between publication and the introduction of the accompanying ideas into the classroom. Practitioners were almost wholly preoccupied with the most recent articles (see Figure 8).

*Number of journals viewed*

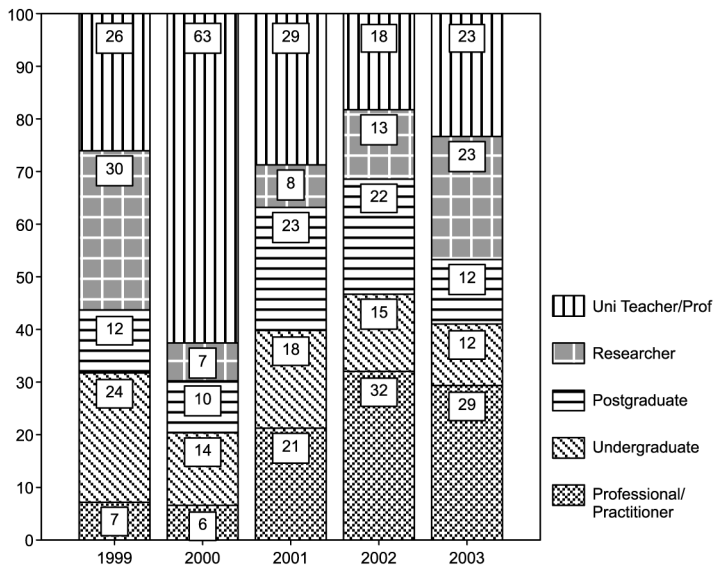
In total, 91 per cent of sessions saw just one journal title being viewed. In fact, undergraduates were most likely to view two or more titles in a session, 35 per cent of sessions had done so. Postgraduates (27 per cent) were next with 27 per cent of sessions viewing two or more titles (see Figure 9).

*Use of added value services – the profile function*

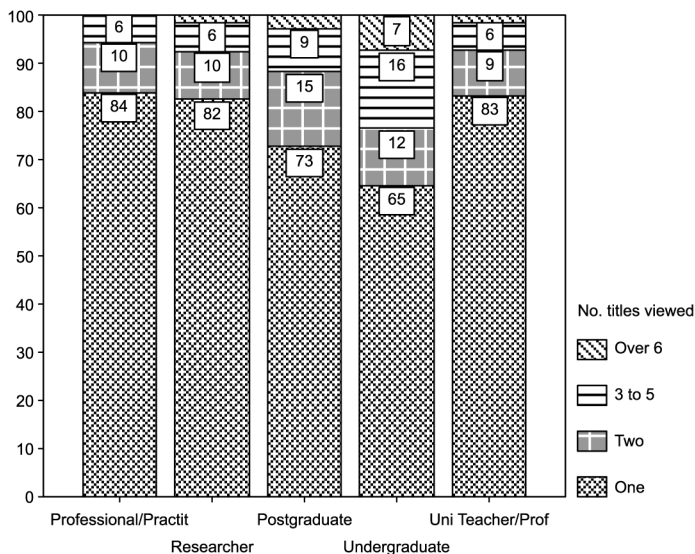
It was learnt from the literature review that students liked to search simply and this analysis tests this hypothesis as it looks at use of the profiling function on Synergy, which is really an advanced (current awareness) function, a form of personalization, the use of which says something about the perceived need to keep up to date. Figure 10 shows that professional/practitioners and postgraduates were more likely to sign-up for a profile, 25 per cent and 22 per cent did so, and professors/teachers were least likely to, only 13 per cent had used the function. Undergraduates were the second least likely to use it, something which supports the hypothesis.

**OhioLINK**

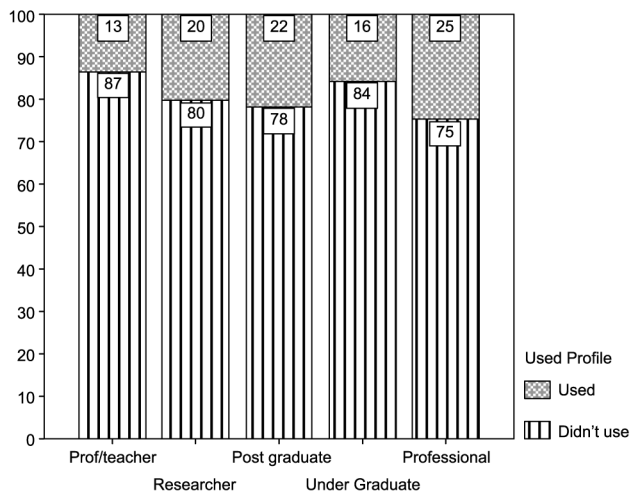
Sub-network computer labels provided information in regard to whether the user was likely to be a student or staff member. Server logs record the connecting computer's IP number. All IP numbers have the same format that is four sets of numbers separated



**Figure 8.**  
Age of article viewed by  
occupational status



**Figure 9.**  
Number of journals  
viewed in a session by  
occupational status



**Figure 10.**  
Use of the profile function  
by occupational status  
of user

by three periods (129.22.7.22). IP numbers at a particular moment in time will identify the connecting network, the sub network and the node. Where the network is the overall organisation, the sub network is a network of computers within the organisation and the node is a computer in the network. IP addresses are interpreted via a process of reverse DNS (Domain Name Server) lookup. This process converts the IP number into named details of the network, sub network and node. For example the IP number 128.40.156.245 translates to the DNS address chemc245.chem.ucl.ac.uk where chemc245 relates to the computer or a number allocated to a computer; chem

identifies the sub network of computers and is taken to be a network with in chemistry and ucl.ac.uk identifies the academic institution UCL based in UK.

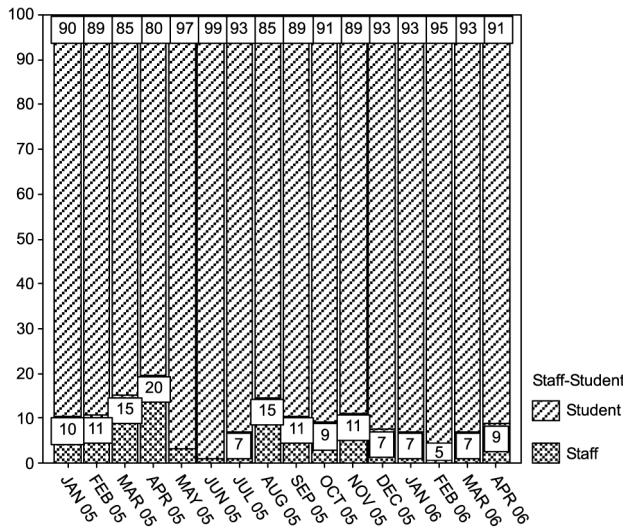
IP numbers were analysed and a reverse DNS lookup was carried out. Five groupings were extracted from the DNS information; the online node name allocated to the computer, the sub network name, the host organisation name, the organisation type and the organisation location. On inspection of the data relating to participating OhioLINK universities it was found that the sub network names for one major university could be accurately identified, especially in regard to a student network and a staff network. The university studied here has more than 2,500 full-time faculty members, slightly less than 3,000 full-time members of staff and about 9,800 students of which about 4,000 are undergraduates.

There are disadvantages in relying on information derived from sub network labels in that this information is generally cryptic or in anagram form and, furthermore, network administrators may select labels that may have little bearing to the physical location of the computer. However, there are compelling reasons why administrators will use meaningful location labels as to do so will provide an easy reminder of the status and location of the computer for the organisation.

*Levels of activity*

There were 5,067 users associated with the staff sub network and about 10 times that number, 48,267, with the student network. This confirms the Synergy finding that students were the majority users of digital resources, there are after all many more of them.

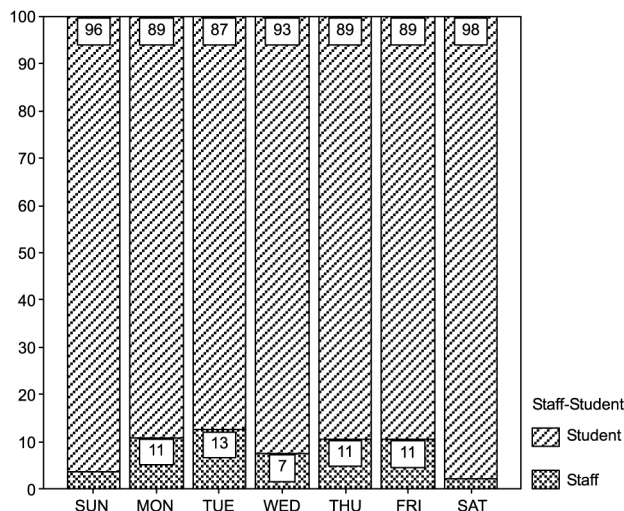
Figure 11 gives the percentage share of page views by sub network label (staff and student) by month and year. June saw students dominate usage, while April saw staff making their most significant contribution, when they accounted for a 20 per cent share of use. Student use is very much tied to the module being studied and these last only ten weeks or so (see Figure 12).



**Figure 11.**  
Percentage distribution of  
page views for staff and  
student by month  
and year



**Figure 12.**  
Percentage share of page views for staff and student by day of week



In terms of day of week, staff accounted for a lower percentage of page views at weekends compared to students. Staff use accounted for about 11 to 13 per cent of total staff and student use during weekdays, except Wednesdays; however, this share fell to about 2 to 3 per cent at weekends. Note these figures are for computers based in the university.

#### *Site penetration*

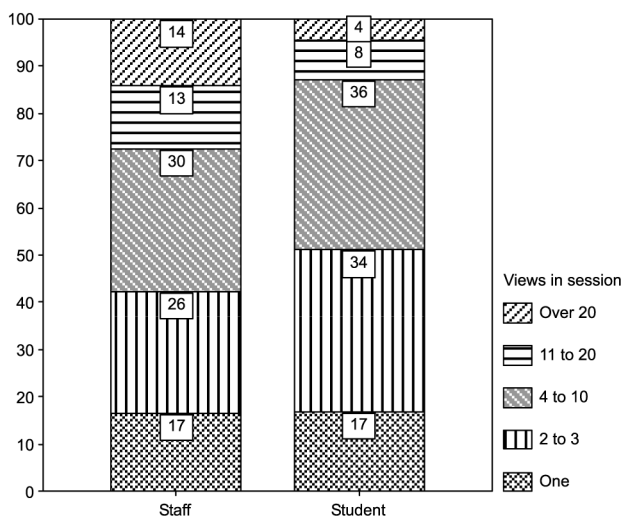
In terms of the number of page views in a session, the staff labelled sub network was identified as making a greater number views in a session than the student labelled network: 27 per cent of sessions saw eleven or more pages viewed in a session; however, just 12 per cent of student sessions viewed that many pages. Students were more likely to view 2 to 3 pages in a session – 34 per cent did so as compared to 26 per cent for staff. This provides support and powerful triangulation for the Synergy data, which showed that students were less likely to penetrate a website deeply (see Figure 13).

#### *Time online*

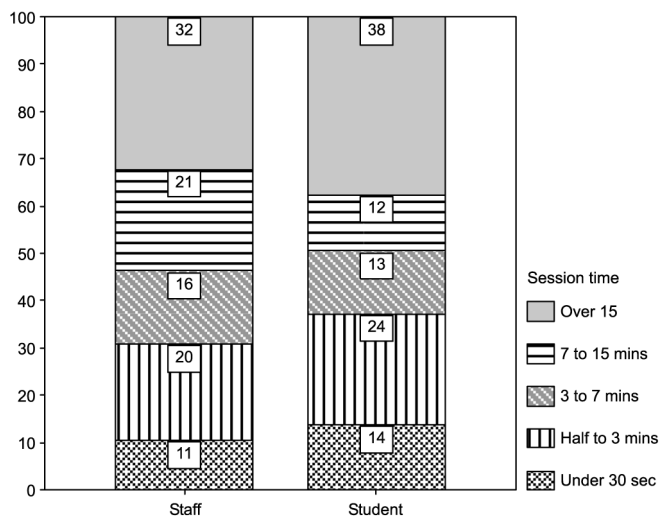
Figure 14 looks at the session time (grouped) for staff and students. Staff were more likely to undertake longer sessions, 69 per cent of their sessions lasted more than 3 minutes compared to 62 per cent of student sessions that lasted this length. However, students were more likely to undertake very long (possibly, reading) sessions lasting more than 15 minutes, 38 per cent as compared to 32 per cent for staff.

There was further evidence to support the belief that students were more likely to read an article online. While 9 per cent of staff computers recorded a read time of between seven and half and one and half hours (the time band that accompanying questionnaire work identified as the time people said it took to read an article online) this was true of 14 per cent of student computers.

In terms of average (median) page view time, broken down for articles and abstracts, staff viewed articles in a much shorter time, 77 seconds compared to 106



**Figure 13.**  
Percentage share of pages  
viewed in a session by  
staff and students



**Figure 14.**  
Percentage share of  
session times by staff  
and students

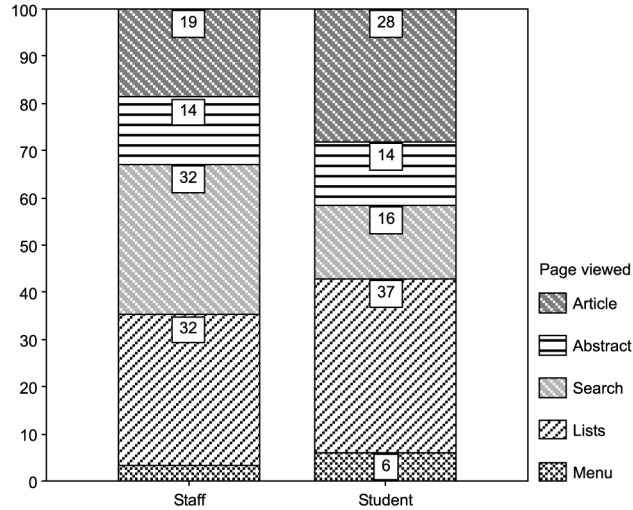
seconds for students, but spent some 3 seconds longer viewing abstracts – 23 compared to 20 seconds.

#### *Type of page viewed*

Figure 15 shows that users on the staff network were more likely to employ the search facility: 32 per cent did so as compared to 16 per cent on the student network, quite a large difference. Staff were less likely to view articles (19 per cent compared to 28 per cent for students) and menus or lists (34 per cent compared to 43 per cent for students).

Concentrating on just article and abstract page views, staff were more likely to view an abstract as compared to students. For staff the split between abstracts and articles

**Figure 15.**  
Percentage share of type of page viewed for staff and students

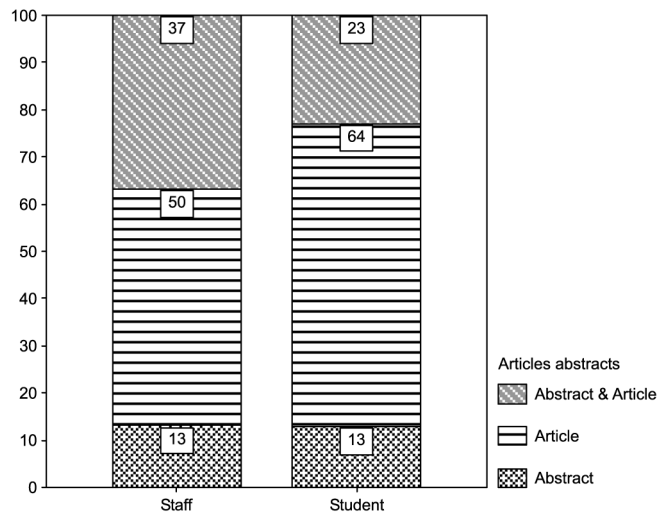


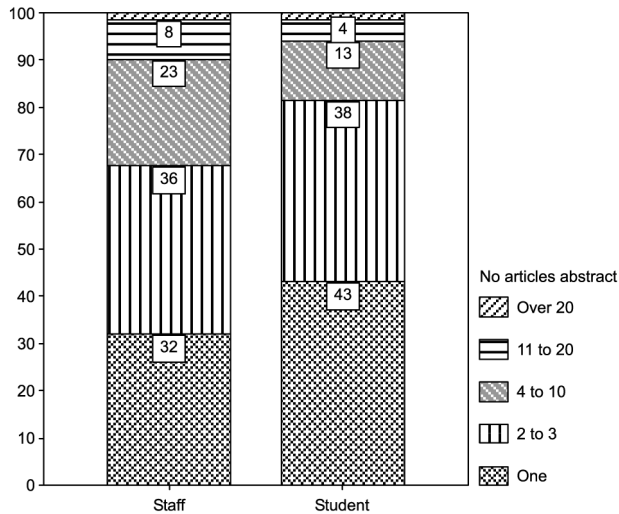
was 43 per cent and 57 per cent compared to 33 per cent and 77 per cent for students. This is another area that merits further qualitative investigation in order to find out whether students are really reluctant to use abstracts and, if so, why.

Figure 16 gives the same distribution but for sessions as clearly users could view both an abstract and article in a session. This demonstrates that in sessions where abstracts were viewed they were often viewed in sessions that also included views to articles; this was true of 37 per cent of staff sessions and 23 per cent of student sessions. Clearly there appears to be a decision process in the selection of material being made here.

Figure 17 gives the share of the number of articles and abstracts viewed in a session by students and staff. Staff were more likely to view a greater number of abstracts and

**Figure 16.**  
Percentage share of sessions distribution of abstracts and article views by staff and students



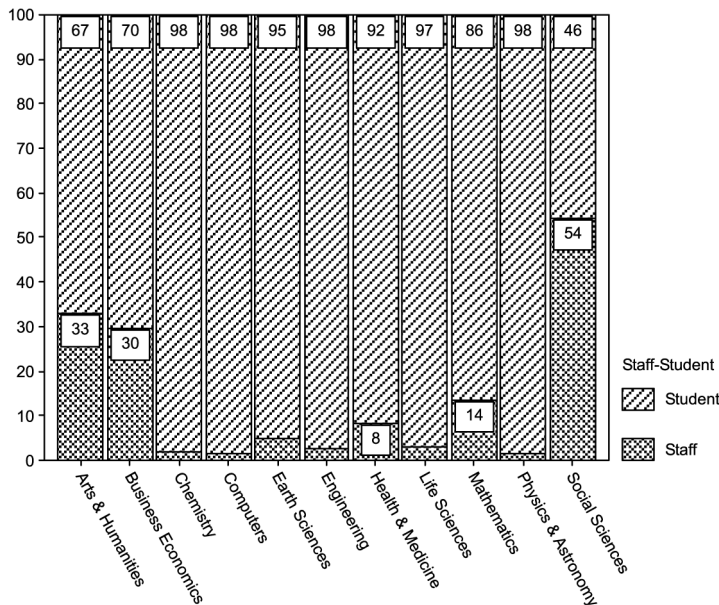


**Figure 17.** Percentage share of the number of articles and abstracts viewed in a session broken down by staff and students

articles in a single session; a third (32 per cent) viewed 4 or more abstracts or articles compared to just 18 per cent of students who accessed this many abstracts or articles in a single session.

*Subject of journals viewed*

Figure 18 gives the breakdown of use for staff and students by the subject of article viewed. There are large differences here with staff accounting for a higher proportion



**Figure 18.** Percentage share of page views for staff and students by subject of the journal

of Social Science use (54 per cent) and a very low proportion of Science use, while students accounted for a lower proportion of Social Science use and a very high proportion of the use in sciences (such as Chemistry and Life Sciences). This is an issue that needs further qualitative investigation. It might be the case that students in Social Sciences and Arts and Humanities rely more on books while students in sciences tend to rely more on journal articles.

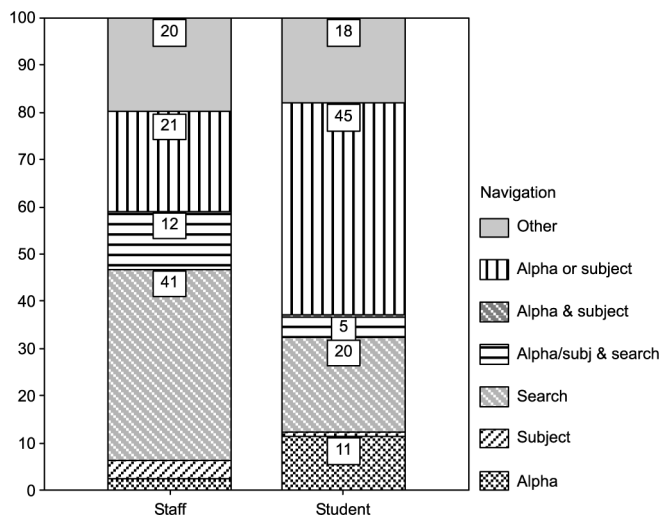
*Method of navigation (Browsers v. searchers)*

A greater proportion of staff sessions only employed the search facility (41 per cent as compared to 20 per cent for students). Staff were also more likely to use other unknown access methods (20 per cent compared to 18 per cent for students). These unknown methods are hypothesised to be direct links maybe by email, copy and pasting into the browser window, RSS links or reference links. Students were more likely to have used the alphabetical and/or subject menus, 61 per cent did so while the equivalent figure for staff was 35 per cent. Interesting findings these, which show that it is not only students who demonstrate a preference for the (internal) search engine, staff appreciate their qualities as well (see Figure 19).

Staff used the search facility more often and they also tended to use the advance search procedure more often too. While nearly two-thirds (65 per cent) of staff searches used advance search only in a session this was true of just over half (54 per cent) of searches conducted by students. Students were more likely to undertake a search using only a simple search: 40 per cent did as compared to 24 per cent of staff.

*Number of journals viewed*

The use of the search facility leads to a wider range of material being viewed and this proved to be the case with staff viewing a greater number of different journals in a session. Well over half (56 per cent) of sessions conducted by staff viewed two or more journals and this compared to 43 per cent for students. Furthermore, a third of staff sessions viewed 4 or more different journals compared with just 14 per cent for student.



**Figure 19.**  
Percentage distribution of sessions by method of navigation to content by staff and students

*Age of articles viewed*

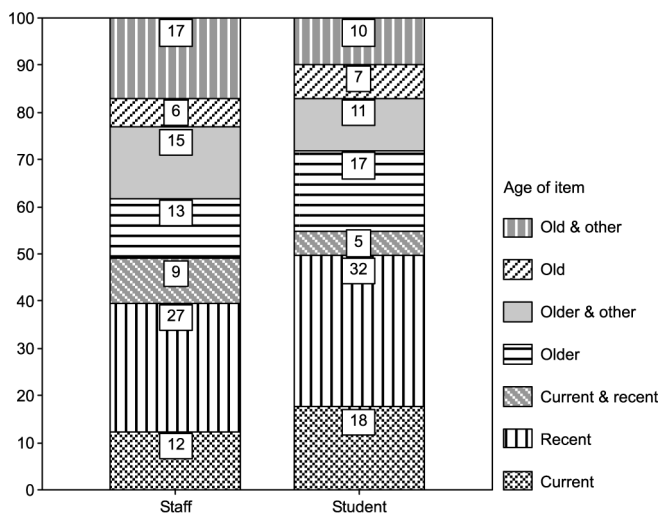
There were differences in the age of articles viewed at the session level. Current in Figure 20 refers to articles of up to one year old; recent articles were 1-3 years old; older articles 4-7 years and old articles more than 7 years old. Staff were less likely to conduct sessions where just current and recent material were viewed. Perhaps, for staff there was a clearer demarcation between research sessions and current material updating. Students clearly viewed much more current material and were more focussed perhaps in their information seeking. This seems to conflict with the findings of Synergy study which found that undergraduate students tended to view older material. However, unfortunately it was not possible to separate undergraduates from postgraduates in the OhioLINK study and this might well account for the difference.

**Oxford Scholarship Online (OSO)**

At this point the paper switches to an examination of e-book use, specifically e-monographs. Sub-network analysis was used to make comparisons between staff and students and again, it was thought to be reasonable to argue that computers located in University College London (UCL) faculties were predominantly used by staff and research students, while computers located in the halls of residence were used by students. UCL has more than 4,000 academic and research staff and about 19,000 students of which more than a third are at graduate level. The following section compares the information seeking behaviour of these two groups in regard to OUP's e-book collection.

*Levels of activity*

A third of UCL usage related to the student halls of residence network, which considering that the Oxford books were monographs (thought to be more suitable for staff) rather than text books, showed a strong interest in e-books amongst students,



**Figure 20.** Share of sessions by mix of age of article viewed in a session broken down by students and staff



something which conflicts with the findings of others (Anuradha and Usha, 2006). In fact as much as sixty per cent of Economics and Finance views were resolved to the student halls of residence.

Students (halls of residence) recorded longer sessions and 47 per cent of sessions were recorded as lasting over fifteen minutes.

*Site penetration*

In terms of page views in a session staff users were both more likely to view just one page, perhaps to see what the book was about, and were less likely just to look at 2 to 3 pages, only 21 per cent did so as compared to users from the halls of residence (see Figure 21).

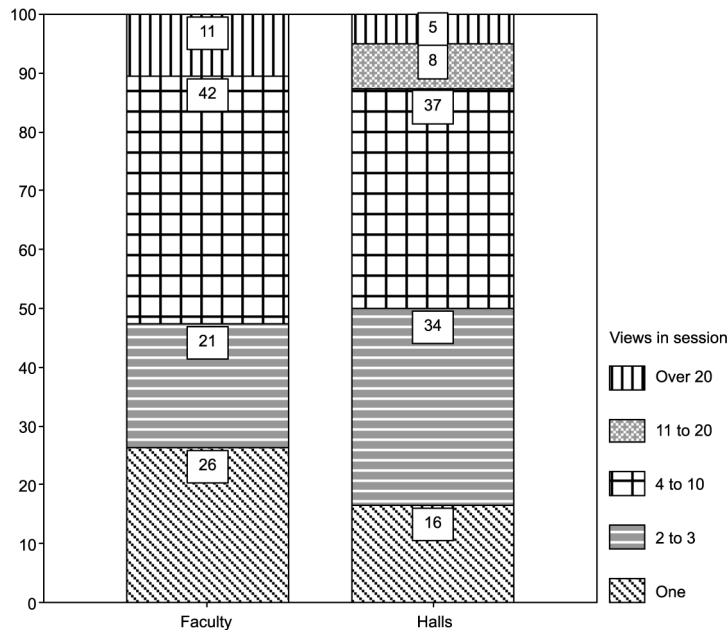
In terms of types of pages viewed staff computers recorded proportionally many more author searches being undertaken and there were fewer views to the site's homepage (see Figure 22).

In terms of page views and print-outs (employing the service's own print facility) made in a session – possible outcome (satisfaction) metrics, staff were more likely to have a page printed, 21 per cent did so as compared to 5 per cent for students. Of course, printing was more likely to be free at the faculty location (see Figure 23).

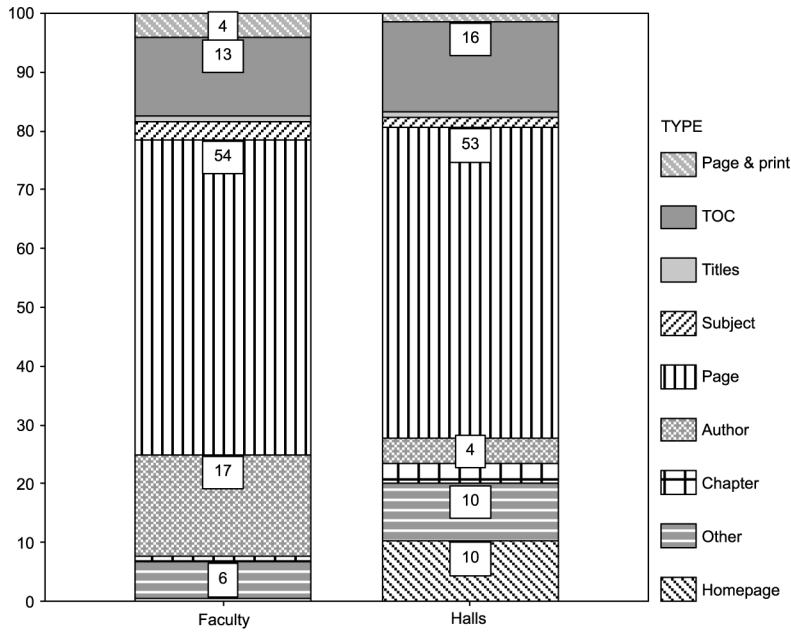
Finally, students were more likely to find OSO titles via the Library catalogue, 65 per cent of those located at the halls of residence used the UCL catalogue. This compared to about 25 per cent for the staff.

**Conclusions and discussion**

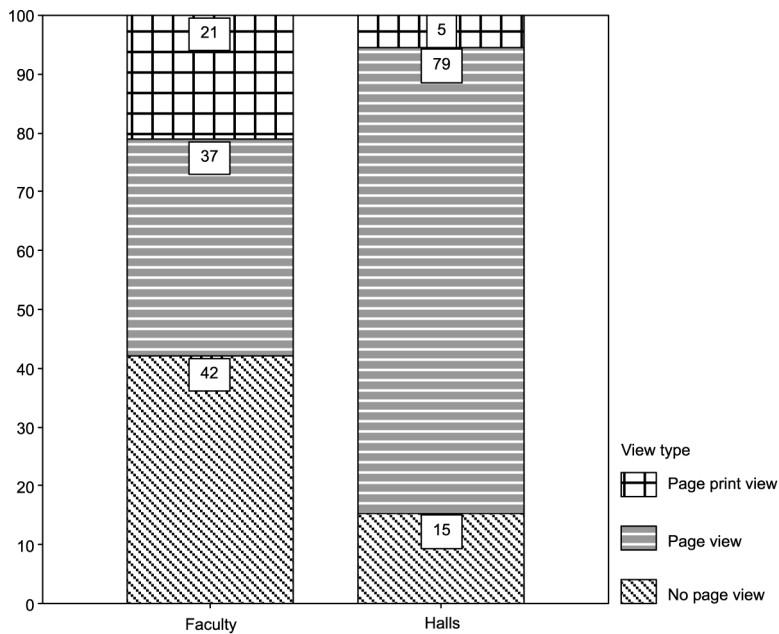
The log analyses presented have generally uncovered quite a distinctive form of information seeking behaviour associated with students and quite appreciable



**Figure 21.**  
Staff v. student  
distribution of number of  
pages viewed in a session



**Figure 22.**  
Type of page viewed for staff V.students



**Figure 23.**  
Staff V.students page views by whether document printed

differences between them and other members of the academic community, and the scholarly user community outside academe. These traits and differences might point to generational changes in information seeking, although, of course, they may be simply or partly a function of just being a student; see also the 'Study on the Information Behaviour of the Researcher of the Future'.

The main findings relate to:

- *Usage.* Students constituted the biggest users in terms of sessions undertaken and pages viewed. The Synergy study showed that, overall, usage declined as academic status increased, which is unsurprising, given the greater networking opportunities available to senior staff and the smaller numbers of people involved. This confirms the TULIP findings which showed that graduate students searched electronic journals more actively and with a broader focus than faculty (Borghuis, 1996). Also a survey commissioned by OCLC (2002) showed that full-text electronic journals were the most used web-based library resource by students and Junni's (2007) examination of references of Masters dissertations showed increasing use of scholarly articles by graduate students. However, the Synergy study also showed that in terms of heavy use (viewing 21 or more pages in a session), the likelihood of being a heavy user actually increased with academic status. Thus the usage profile of undergraduates is that they conduct many sessions but do not view a lot of pages during a session. This may be a consequence of their heavier use of internet wide search engines. This all fits the picture of students as "bouncers" established by the authors (Nicholas *et al.*, 2007). However, this turned out not to be the case with e-books where students viewed more pages in a session than staff. This could be because e-books are a more appropriate form of e-resource to students, which seems logical. This is something that will be further examined in the current study.
- *Type of page viewed.* The picture here is less straightforward. Thus take abstract viewing as an example. In the case of Synergy, surprisingly perhaps, undergraduates proved to be the biggest users of abstracts, and this was also the finding of the Borghuis (1996) study. However, in the case of OhioLINK, where students had a genuine level playing field regarding what they viewed and complete freedom to view full-text, staff were more likely to view an abstract as compared to students. Part of the explanation might be that the OhioLINK study incorporates postgraduates in the student figure. In regard to full-text viewing, the Synergy study showed that the use of PDFs increased as users moved up the academic scale, from undergraduate to professor/teacher.
- *Searching and navigating.* Undergraduates and postgraduates were the most likely users of library links to access scholarly databases, suggesting an important "hot link" role for libraries. In terms of the browsing V.searching issue the picture is somewhat confused, with undergraduates the most likely to use the search facility in the case of Synergy but more likely to use the alphabetical or subject menu in the case of OhioLINK. This could be due to differences in the design and content of the databases and points to the danger of making general assertions about undergraduates and their search engine pre-occupation. When searching students were however more likely to undertake a simple search, as

---

the OhioLINK study demonstrated. In the case of e-books staff exhibited a far greater preference for the author search than students.

- *Reading online.* The OhioLINK study showed that students were more likely to record long online sessions lasting more than 15 minutes, evidence, perhaps, of substantial online reading, which is borne out by the project's associated questionnaire data (Nicholas *et al.* 2008). Students were much more likely to read online than other academic groups and this was partly to do with personal preferences and partly to do with the print charges students are faced with in many institutions. This finding is supported by a survey conducted by Outsell Inc for the Digital Library Federation and the Council of Library and Information Resources (Friedlander, 2002) which found that undergraduates were more willing to rely on electronic resources than graduates and faculty, with approximately half using electronic resources exclusively or almost exclusively. A survey carried out at the University of Strathclyde (Abdullah and Gibb, 2006), which investigated the awareness of e-books amongst students, found that the majority of users (94 per cent) read them on-screen.
- *Subject diversity.* The OhioLINK study showed that there were big differences here between staff and students, with staff accounting for a higher proportion of Social Science use but a much lower proportion of Science use. This diversity finding is born out by others (Friedlander, 2002).
- *Currency.* In the case of OhioLINK students clearly viewed much more current material but in the case of Synergy the opposite was true, with undergraduates showing a preference for older articles. The fact that it was not possible to separate undergraduates from postgraduates in the OhioLINK study might well account for the difference.
- *Number of journals viewed.* Again the findings are at odds, with OhioLINK showing that staff viewed more journals in a session than students and Synergy showing that the opposite was true.

Overall, in regard to e-resource use, the research literature tends to concur (unsurprisingly) that it is on the increase and there is a reliance on simple searching, and students get better at searching as their skills as they progress to the higher stages of their studies. We also know from these studies that many factors influence their information seeking, and academics play an important role in shaping it. And this might well explain the contradictions and anomalies we have found. In regard to this study two factors would appear to be particularly significant:

- (1) *platform differences* – the site design and functionality differs; and
- (2) *cultural factors* – OhioLINK is an exclusively American service, OSO was studied exclusively in a UK context and Synergy has an international audience.

While the differences in information seeking behaviour between scholarly communities has been highlighted it would be a mistake to believe that it is only students' information seeking that has been fundamentally shaped by huge digital choice, easy (24/7) access to scholarly material, disintermediation, and very powerful and influential search engines. The same, of course, has happened to professors, lecturers and practitioners. Virtual Scholar research has shown that a considerable

number of users exhibit a bouncing/flicking behaviour, which sees searching conducted horizontally, rather than vertically. Power browsing and viewing appear to be the norm for many; reading appears to be undertaken only occasionally online, probably undertaken offline and in some cases not done at all.

It needs to be borne in mind that the Virtual Scholar studies that have been evaluated were of users of specific digital services and were based on records of what people did and not what they say they did. Virtually all of the studies that have been identified in the literature review were based on surveys and interview studies and clearly there is problem of recall and a strong likelihood that students (and staff) are going to tell librarians and researchers what they think they want to hear. These studies also tend to cover all types of information seeking activities of students (catalogue, library use etc) as well as the use of e-resources. They also tend to cover non-users, however given the huge popularity of such services it is highly unlikely that students constitute a large population of non-users. Therefore, these methods (log analysis, surveys and interviews) should be used in conjunction to build a clear picture of students' information-seeking behaviour and to provide an explanation for the observed behaviour.

As to the future, the arrival of social networking sites and the popularity of blogs (provided, for instance on the Intute site) will undoubtedly further shape student information seeking behaviour. However, studies of Intute and the BL Learning site conducted as part of the GoogleGeneration study show that student behaviour is still essentially traditional in nature (CIBER, 2007).

### References

- Abdullah, N. and Gibb, F. (2006), "A survey of e-book awareness and usage amongst students in an academic library", *Proceedings of International Conference of Multidisciplinary Information Sciences and Technologies, Merida, Spain, 25-28 October, 2006*, available at: [http://eprints.cdlr.strath.ac.uk/2280/01/FGibb\\_survey\\_ebook.pdf](http://eprints.cdlr.strath.ac.uk/2280/01/FGibb_survey_ebook.pdf)
- Anuradha, K.T. and Usha, H.S. (2006), "Use of e-books in an academic and research environment: a case study from the Indian Institute of Science", *Program: electronic library and information systems*, Vol. 40 No. 1, pp. 48-62.
- Barrett, A. (2005), "The information-seeking habits of graduate student researchers in the humanities", *Journal of Academic Librarianship*, Vol. 31 No. 4, pp. 324-31.
- BBC News* (2006), "Net student think copying OK", 18 June, available at: <http://news.bbc.co.uk/go/pr/fr/-/1/hi/education/5093286.htm>
- Becker, N.J. (2003), "Google in perspective: understanding and enhancing student search skills", *New Review of Academic Librarianship*, Vol. 9 No. 1, pp. 84-100.
- Bilal, D. (1998), "Children's search processes in using World Wide Web search engines: an exploratory study", *Proceedings of the ASIS Annual Meeting*, Vol. 35, pp. 45-53.
- Bilal, D. (2000), "Children's use of the Yahoo!igans! Web search engine: 1. Cognitive, physical, and affective behaviors on fact-based search tasks", *Journal of the American Society for Information Science*, Vol. 51 No. 7, pp. 646-65.
- Borghuis, M. (1996), *TULIP Final Report*, Elsevier Science, New York, NY, available at: [www.elsevier.com/wps/find/librariansinfo.librarians/tulipfr](http://www.elsevier.com/wps/find/librariansinfo.librarians/tulipfr) (accessed 11 July 2007).
- Brabazon, T. (2007), *The University of Google: Education in the (Post) Information Age*, Ashgate, London.

- 
- CIBER (2007), *Information Behaviour of the Researcher of the Future ("Google Generation" Project)*, CIBER, University College London, available at: [www.ucl.ac.uk/slais/research/ciber/downloads/](http://www.ucl.ac.uk/slais/research/ciber/downloads/) (accessed 17 March 2008).
- Dalgleish, A. and Hall, R. (2000), "Uses and perceptions of the World Wide Web in an information-seeking environment", *Journal of Librarianship and Information Science*, Vol. 32 No. 3, pp. 104-16.
- Drabenstott, K.M. (2003), "Do nondomain experts enlist the strategies of domain experts?", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 9, pp. 836-54.
- Entwistle, N. (2003), *Concepts and Conceptual Frameworks Underpinning the ETL Project, Occasional Report 3*, ETL (Enhancing Teaching-Learning Environments in Undergraduate Courses) Project, University of Edinburgh, Edinburgh.
- Eskola, E. (2005), "Information literacy of medical students studying in the problem-based and traditional curriculum", *Information Research*, Vol. 10 No. 2, available at: <http://informationr.net/ir/10-2/paper221.html> (accessed 11 July 2007).
- Fast, K.V. and Campbell, G. (2004), "I still like Google: university student perceptions of searching OPACs and the Web", *Proceeding of the 67th ASIS&T Annual Meeting*, Vol. 41, pp. 138-46.
- Fidel, R., Davies, R.K., Douglass, M.H., Holder, J.K., Hopkins, C.J., Kushner, E.J., Miyagishima, B.K. and Toney, C.D. (1999), "A visit to the information mall: web searching behavior of high school students", *Journal of the American Society for Information Science*, Vol. 50, pp. 24-37.
- Friedlander, A. (2002), "Dimensions and use of the scholarly information environment", Digital Library Federation and Council on Library and Information Resources, Washington, DC, available at: [www.clir.org/PUBS/reports/pub110/contents.html](http://www.clir.org/PUBS/reports/pub110/contents.html) (accessed 11 July 2007).
- Graham, L. and Metaxas, P.T. (2003), "Of course it's true; I saw it on the internet: critical thinking in the internet era", *Communication of the ACM*, Vol. 46 No. 5, pp. 71-5.
- Heinström, J. (2005), "Fast surfing, broad scanning, and deep diving: the influence of personality and study approach on students' information-seeking behaviour", *Journal of Documentation*, Vol. 61 No. 2, pp. 228-47.
- Jamali, H.R., Nicholas, D. and Huntington, P. (2005), "Use and users of scholarly e-journals: a review of log analysis studies", *Aslib Proceedings*, Vol. 57 No. 6, pp. 554-71.
- Junni, P. (2007), "Students seeking information for their Masters' theses: the effect of the Internet", *Information Research*, Vol. 12 No. 2, available at: <http://InformationR.net/ir/12-2/paper305.html> (accessed 17 March 2008).
- Kerins, G., Madden, R. and Fulton, C. (2004), "Information seeking and students studying for professional careers: the case of engineering and law students in Ireland", *Information Research*, Vol. 10 No. 1, available at: <http://informationr.net/ir/10-1/paper208.html> (accessed 11 July 2007).
- Liu, Z. and Yang, Z.Y. (2004), "Factors affecting distance-education graduate students' use of information sources: a user study", *Journal of Academic Librarianship*, Vol. 30 No. 1, pp. 24-35.
- Nicholas, D., Huntington, P., Jamali, H.R. and Dobrowolski, T. (2007), "Characterising and evaluating information seeking behaviour in a digital environment: spotlight on the 'bouncer'", *Information Processing and Management*, Vol. 43 No. 4, pp. 1085-102.
- Nicholas, D., Huntington, P., Jamali, H.R. and Watkinson, A. (2006), "The information seeking behaviour of the users of digital scholarly journals", *Information Processing and Management*, Vol. 42 No. 5, pp. 1345-65.



- Nicholas, D., Huntington, P., Lievsley, N. and Wasti, A. (2000), "Evaluating consumer web site logs: case study *The Times/Sunday Times* web site", *Journal of Information Science*, Vol. 26 No. 6, pp. 399-411.
- Nicholas, D., Huntington, P., Tenopir, C., Jamali, H.R., Dobrowolski, T. and Rowlands, I. (2008), "An appraisal of the full-text download, the gold standard usage metric", *Aslib Proceedings*, Vol. 60 No. 3, pp. 186-98.
- OCLC: Online Computer Library Center (2002), *How Academic Librarians can Influence Students' Web-based Information Choices*, OCLC White Paper on the Information Habits of College Students, Online Computer Library Center, Dublin, OH, available at: [www5.oclc.org/downloads/community/informationhabits.pdf](http://www5.oclc.org/downloads/community/informationhabits.pdf) (accessed 2 December 2005).
- Prabha, C., Connaway, L.S., Olszewski, L. and Jenkins, L.R. (2007), "What is enough? Satisficing information needs", *Journal of Documentation*, Vol. 63 No. 1, pp. 74-89.
- Rowley, J. and Urquhart, C. (2007), "Understanding student information behaviour in relation to electronic information services: lessons from longitudinal monitoring and evaluation, Part 1", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 8, pp. 1162-74.
- Schacter, J., Chung, G.K.W.K. and Dorr, A. (1998), "Children's internet searching on complex problems: performance and process analyses", *Journal of the American Society for Information Science*, Vol. 49 No. 9, pp. 840-9.
- Simon, H.A. (1979), *Models of Thought*, Yale University Press, New Haven, CT.
- Urquhart, C. and Rowley, J. (2007), "Understanding student information behaviour in relation to electronic information services: lessons from longitudinal monitoring and evaluation, Part 2", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 8, pp. 1188-97.
- Urquhart, C., Thomas, R., Lonsdale, R., Spink, S., Yeoman, A., Fenton, R. and Armstrong, C. (2003), "Uptake and use of electronic information services: trends in UK higher education from the JUSTEIS project", *Program*, Vol. 37 No. 3, pp. 168-80.
- Valentine, B. (1993), "Undergraduate research behaviour: using focus groups to generate theory", *Journal of Academic Librarianship*, Vol. 19 No. 5, pp. 300-4.
- Whitmire, E. (2002), "Disciplinary differences and undergraduates' information-seeking behaviour", *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 8, pp. 631-8.
- Whitmire, E. (2003), "Epistemological beliefs and the information-seeking behavior of undergraduates", *Library and Information Science Research*, Vol. 25 No. 2, pp. 127-42.

**Corresponding author**

David Nicholas can be contacted at: [david.nicholas@ucl.ac.uk](mailto:david.nicholas@ucl.ac.uk)