



Does Library Use Affect Student Attainment? A Preliminary Report on the Library Impact Data Project

Graham Stone, Dave Pattern, Bryony Ramsden

University of Huddersfield, UK,
G.Stone@hud.ac.uk

Abstract

The current economic climate is placing pressure on UK Universities to maximise use of their resources and ensure value for money. In parallel, there is a continuing focus on the student experience and a desire that all students should achieve their full potential whilst studying at University.

Internal investigation at the University of Huddersfield suggests a strong correlation between library usage and degree results, and also significant under-usage of expensive library resources at both school and course level. Data from over 700 courses using three indicators of library usage; access to e-resources; book loans and access to the library were matched against the student record system and anonymised.

Initial findings highlighted that the correlation between library usage and grade had not yet been significance tested. In January 2011, the University of Huddersfield, together with partners at the Universities of Bradford; De Montfort; Exeter; Lincoln; Liverpool John Moores; Salford and Teesside were awarded JISC funding to prove the hypothesis that *there is a statistically significant correlation across a number of universities between library activity data and student attainment.*

Academic librarians at Huddersfield are also working closely with tutors on a selected sample of courses to explore the reasons for unexpectedly low use of library resources. By identifying subject areas or courses which exhibit low usage of library resources, service improvements can be targeted such as:

- course profiling, to determine the particular attributes of each course and its students, which may affect library use
- targeted promotion of resources at the point of need

- raising tutor awareness of resources, particularly e-resources and current awareness services
- review of the induction process
- target information resources allocation, to ensure value for money
- target staffing resources, to ensure that support for students is available at key times of the year.

This paper will report on the initial findings of the project and whether the measurable targets have been achieved:

- Sufficient data are successfully captured from all partners
- Statistical significance is proved for all data
- The hypothesis is either wholly or partly proved for each data type and partner

Key Words: Library usage; student attainment; low use; non-use; academic libraries; undergraduate students; achievement

Introduction

In 2010, the University of Huddersfield reported on its analysis of anonymised library usage data (access to e-resources, book loans and access to the library against student attainment) (White and Stone, 2010) from over 700 courses over four years (2005/6–2008/9) against student attainment. At the time it was suggested that there appeared to be a strong correlation between usage data and student attainment at both school and course level, although this had yet to be proved to be statistically significant.

The work coincided with the recent Comprehensive Public Spending Review and the Lord Browne's Review of Higher Education Funding and Student Finance. These reports, combined with the continuing focus on the student experience and a desire that all students should achieve their full potential whilst studying at University led to the University of Huddersfield along with seven partner institutions bidding for JISC funding as part of the Activity Data programme, where potential bidders were asked to put forward a hypothesis as part of their project proposal.

This paper will describe the remit of the Library Impact Data Project (LIDP) and outline the methodology used in analysing data from the project partners. It will then go on to discuss initial findings, focus groups and paral-

lel work surrounding non/low use of library resources being undertaken at the University of Huddersfield before highlighting areas of possible further research.

Literature Review

Various studies have attempted to investigate how to measure library performance and its connection with student success. Much of the research conducted was largely at a school library level, particularly in the United States and Canada. In a huge sample (800 elementary schools, 50,000 students, with a sample specifically of grades 3 and 6) the Ontario Library Association (Ontario Library Association, 2006) asked '[d]o school library resources and staff have an impact on students' attitudes towards reading and on their scores on large scale standardized tests?' Using surveys already completed nationally, they found correlations between library staffing and reading performance in both grades, as well as a decline in enjoyment of reading correlating with a decline in staffing of libraries. Similarly in a study of three Ugandan schools with varying levels of library access, Dent (Dent, 2006) found that those students with library access scored higher in particular subjects than those who did not have access. However, overall time spent reading in each student was similar, with those students without library access spending a small amount of time more on reading.

At higher education level, De Jager examined book borrowing in particular. In her 2002 conference paper (De Jager, 2002a), she studied use of short loan stock and 'open shelf' items (i.e. items freely available for loan rather than housed in a separate collection) and found correlations between borrowing and the final passing grade in some courses. However, she felt further investigation was required to look closer at the habits of students achieving particular grades. She took a sample of high-achieving students (70% or above for their final score) from humanities and science courses and focussed specifically on the open shelf collection (De Jager, 2002b). Her findings were surprising: humanities borrowing was at high levels while science students borrowed comparatively little. De Jager accepts that further analysis is required incorporating e-resource usage to paint a broader picture of library use and attainment.

In a paper on the Google Generation and their information-seeking behaviour, Rowlands *et al.* (Rowlands *et al.*, 2008) discuss the need for changing

branding of libraries. Regardless of the image of the Google Generation being highly skilled with searching for online materials and discarding traditional resources, previous research cited by Rowlands *et al.* (OCLC, 2006) demonstrates a continuing desire of students to refer to books, while other studies find an overestimation of the Google Generation's electronic information-seeking skills by students. Gross and Latham (Gross and Latham, 2007) found the lower the skill the students had, the more they overestimated their skills, while Weiler (Weiler, 2005) notes that the tendency to overestimate skills stems from the assumption students know a great deal about the Internet 'as a "cool" medium' (p. 50).

Some research has already been carried out by Huddersfield indicating a relationship between overall library use and attainment (Goodall and Pattern, 2011; White and Stone, 2010). Preliminary work also indicates that e-resource access at a moderate level does not necessarily equate to degree attainment, i.e. at a usage level of 21–40 and 41–60 logins, those achieving first and third degrees had roughly the same number of logins (Pattern, 2010). Clearly there are also other considerations necessary here such as duration of database use, the nature of how they searched, or what they used when they logged.

The Library Impact Data Project

The Library Impact Data Project (LIDP) is a collaborative project between the University of Huddersfield and seven partners: University of Bradford; De Montfort University; University of Exeter; University of Lincoln; Liverpool John Moores University; University of Salford and Teesside University. The project was awarded JISC funding for 6 months (February–July 2011) to prove the hypothesis that '*there is a statistically significant correlation across a number of universities between library activity data and student attainment*'.

It is important to note that the project has acknowledged that the relationship between the two variables is not a causal relationship and there will be other factors which influence student attainment. The project's overall goal is to prove the hypothesis, thereby encouraging greater use of library resources and ultimately to ensure that student attainment is improved particularly in areas of non/low use. This will in turn create tangible benefits to the wider Higher Education (HE) community by creating a better

understanding of the link between library activity data and student attainment. Planned outcomes of the project include the release of the data on an Open Data Commons Licence and a toolkit to allow other HE institutions to benchmark their data.

The project has an active project blog, which is being used to report via a number of themed posts throughout the duration of the project. These include the project plan; the hypothesis; users; benefits; technical and standards; licensing and reuse of software and data; wins and fails (lessons along the way) and a final post written at the end of July.

Legal Issues

A major issue identified at the very beginning of the project was the need to abide by legal regulations and restrictions, such as data protection. The very nature of the data being used in the project makes it sensitive and there is obvious need to ensure complete anonymisation. The team liaised with JISC Legal at the outset of the project and subsequent further discussion with the University of Huddersfield Legal and Data Protection Officers have helped to ensure that there is complete anonymisation.

All partners need to match their usage data to student attainment using an identifier, but once the data have been combined this identifier is removed, thus ensuring anonymity. In order to prevent the identification of individuals at course level, small courses where the cohort is less than 35 students or where fewer than 5 students have obtained a specific degree level have been excluded. The decision as to whether to release the data from all partners as one complete set will be discussed below, if this route is not taken the project will also ensure that no partner can be identified.

Going forward, the plan is to adopt a recommendation from the Using OpenURL Activity Data project in order to notify users of our data collection:

‘When you search for and/or access bibliographic resources such as journal articles, your request may be routed through the UK OpenURL Router Service (openurl.ac.uk), which is administered by EDINA at the University of Edinburgh. The Router service captures and anonymises activity data which are then included in an aggregation of data about use

of bibliographic resources throughout UK Higher Education (UK HE). The aggregation is used as the basis of services for users in UK HE and is made available to the public so that others may use it as the basis of services. The aggregation contains no information that could identify you as an individual.'

Data Issues

The project anticipated that there may be issues in collecting the data. Due to the short timescale of the project, this was seen as a significant risk. All potential partners were asked if they could provide at least two of the three measures of usage as well as the student attainment data (see Table 1), ideally in a machine-readable format such as Excel, XML or CSV.

Table 1: Data Requirements for Project Partners

Data Requirements for Project Partners	
For a specific academic year (e.g. 2009/10), extract details for each graduating student	
<ul style="list-style-type: none"> • academic year of graduation • course title • length of course in years • type of course • grade achieved • school/academic department • number of items borrowed from library <ul style="list-style-type: none"> ◦ either the total number borrowed by that student ◦ or separate values for each academic year • number of visits to the library <ul style="list-style-type: none"> ◦ either the total number of visits by that student ◦ or separate values for each academic year • number of logins to e-resources (or some other measure of e-resource usage) <ul style="list-style-type: none"> ◦ either the total number of logins made by that student ◦ or separate values for each academic year 	e.g. 2009/10 Software Development 3 post grad 2:1 School of Computing 50 items during the 3 years of the course 11 items in 2007/8, 16 in 2008/9 and 23 in 2009/10
If you have other library usage data — e.g. number of library PC logins — please feel to include that in the extract.	

One partner ran into problems at this stage when they found out that although their gate entry system did keep historical data it was stored by the system supplier and was therefore not readily available. This will prove a valuable lesson for future procurement of such systems. In addition, although the attainment data was available for 2010, two-thirds of the identifiers had been deleted as is institutional policy. Lessons were learned and the institution has now put processes in place in order to be able to capture the data from 2011 onwards.

Methodology

At the time of writing, all data had been received by Huddersfield and are currently being processed using SPSS. Some institutions were unable to supply a full set of data for reasons outlined above; in addition some could only supply log-in information, or supplied data in a format that could not be validly compared with other institutions, e.g. book issues and renewals in a combined set. However, these institutions are being analysed as a set of data in their own right, and will be discussed as such in the final report.

Basing an initial analysis of the data on work conducted by David Pattern prior to the project, a non-normal distribution was expected, and it was tested using the Kruskal-Wallis test. A null hypothesis of 'there is no difference between degree results and library usage' was proposed for each type of data: if the null hypothesis can be discarded on the basis of the Kruskal-Wallis test, further analysis can be conducted to confirm where differences lie between degree results. The data sets are large and so it is accepted that the results may be skewed.

The test first asks the data to be checked for distribution using the Kolmogorov-Smirnov Test for normality. Having confirmed that the data does not follow a normal distribution, the Kruskal-Wallis test is run to check for significant differences between groups. The Monte Carlo Estimate was applied to all data, a method of repeatedly testing random samples from a simulated data set mirroring the actual data's distribution to measure the significance: due to the large size of the sample an exact result cannot be calculated. However, the test does not identify where the differences lie, so

further analysis is conducted using the Mann-Whitney U test, which measures differences between selected values. The nature of the Mann-Whitney (and many other tests of difference) means that the more tests conducted for measuring significant differences, the greater the level of significance must be to ensure the test is valid, i.e. testing for a significance of 5% with one test would require significance at 0.05 or lower, but running 5 tests at 5% would require a significance value of 0.01 for each test to prove valid (5% divided by the number of tests conducted). In order to ensure valid significance a maximum number of 3 Mann-Whitney tests were run for each data group, with groups selected on the basis of visual indication from boxplots of the data. Data processing has in some cases shown differences between results and varying types of usage at a significant level, but on examination of the boxplot and removal of lower-level degrees, these have proven to be insignificant. In these cases the data are considered to show no difference between results. Huddersfield's data analysis is shown below as an example.

The University of Huddersfield Data for 2007

Having conducted the Kolmogorov-Smirnov test and found confirmation of non-normality of the e-resource data, the Kruskal-Wallis test provided a highly significant result for difference between values. The box plot in Figure 1 identifies potential differences to be calculated for significance. Points to note for further consideration in later analysis will be outlying usage figures, for example in students achieving a lower second-class degree, extreme outliers are clearly visible, and to a lesser extent in third-class degree access. On the basis of the box plot, an analysis was conducted between first and upper second class, first and lower second class, and first and ordinary degrees. The Mann-Whitney U test found significant differences between first and lower second-class degree access, and between first and ordinary degree access, but not between first and upper second class degree access (which measured at a significance level of $p < 0.08$, and visually appears to be different).

Initial Findings

At this stage it does look like the project will be able to prove that there is a relationship and variance with the data. This implies that what can be seen in Figures 2–4 can be believed and that it can be believed across a range of data and subjects.

Fig. 1: Boxplot of e-resource usage levels by final degree result.

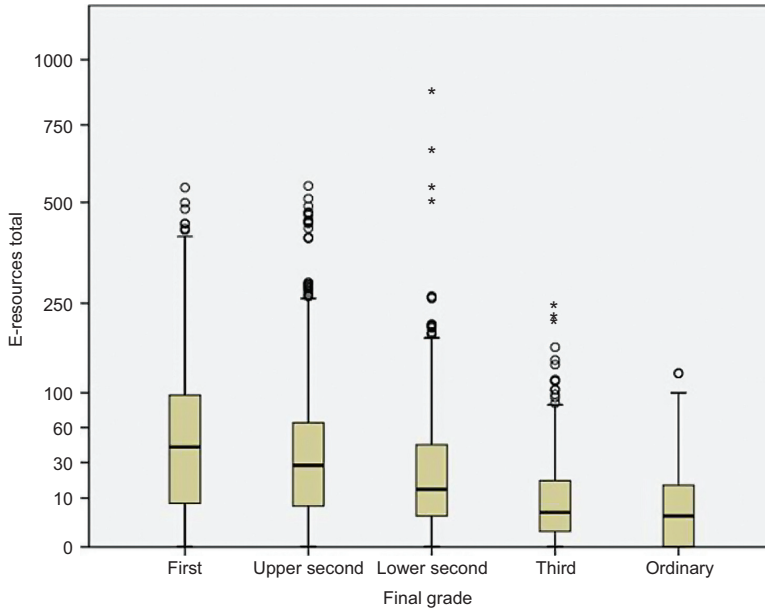


Figure 2 shows the relationship between book loans (including renewals) and student attainment for one of the partners.

Figure 3 shows a similar relationship between book loans and Athens (e-resources authentication) and attainment from another partner.

Despite the apparent correlation between attainment and book loans and attainment and e-resources, data gathered so far seem to show no such correlation between library use or PC logins (see Figure 4).

Information from some of the focus groups (see below) has helped to explain this lack of correlation. The University of Exeter found that although most students use the library regularly, there was a clear division between those students who prefer working in the library and those who prefer working at home. This is likely to have as much to do with personal preference as with engagement with the course. This was backed up by work previously undertaken at Huddersfield that showed that library space was used for more than studying (Ramsden, 2011).

Fig. 2: Relationship between book loans (including renewals) and student attainment.

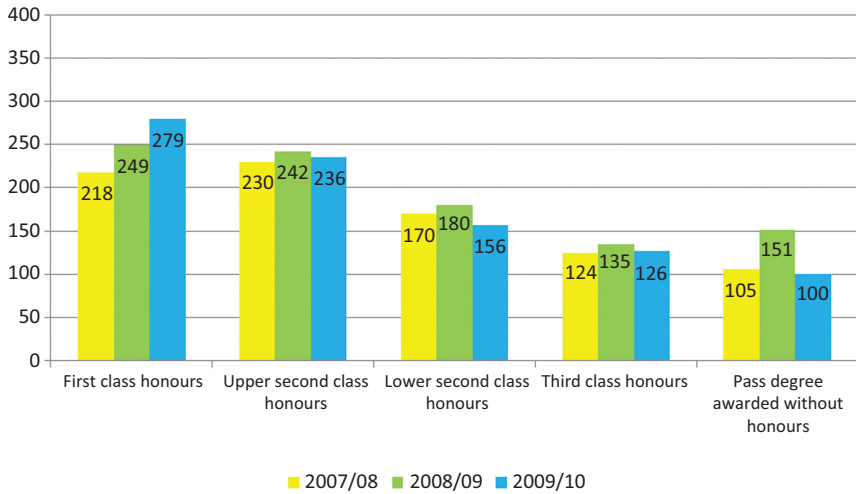


Fig. 3: Relationship between book loans/Athens (e-resources authentication) and student attainment.

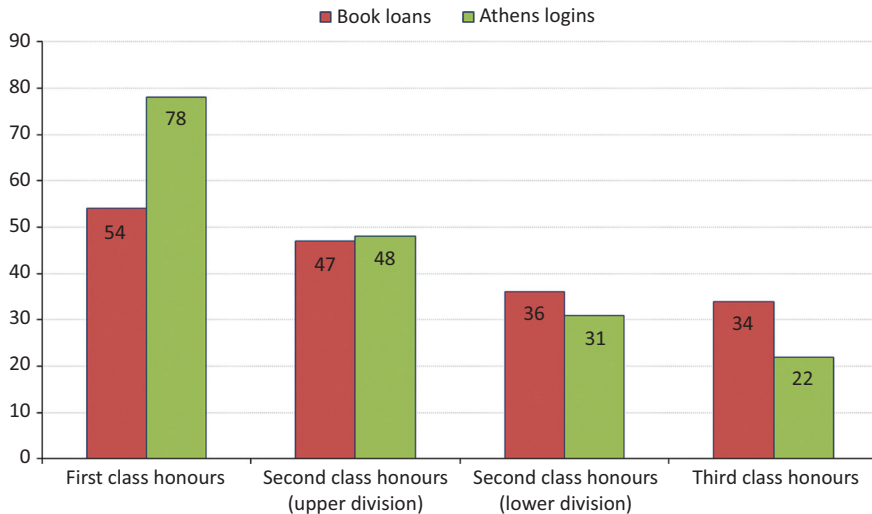
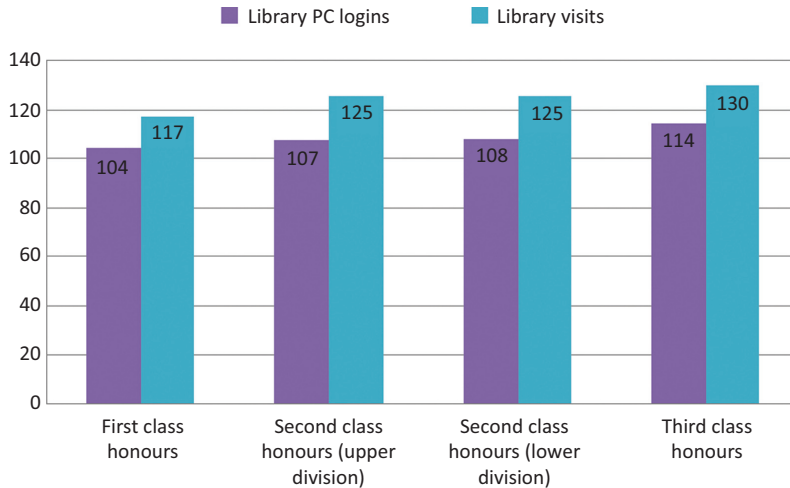


Fig. 4: Relationship between PC logins/library visits and student attainment.



Focus Groups

All participating institutes were asked to conduct focus groups to gather qualitative data on reasons why students may be low or high users of library resources. A set of questions was designed to cover various elements, including:

- resource selection
- frequency of library use (including some discussion of whether they interpreted their usage level as high or low)
- where they accessed e-resources and where they used library resources as a whole, e.g. in the classroom, at home
- had they experienced any difficulties in using or accessing resources?
- had they ever attended any library training sessions or similar?
- how often they read outside of the recommended reading titles?
- what was their experience of libraries like on an educational level prior to attending university?
- whether the library provided a supportive learning environment for their own personal study needs.

The focus group questions were supplemented by a short qualitative survey mirroring some of the themes above, but designed to gather answers

with less bias and avoid any pressure the focus group setting may induce. Additionally, a script introducing the purpose of the group with information about the project, a consent form, and an information sheet covering anonymity issues and contact information were designed for use by each collaborating institution. The project partners were asked to check through the questions and forms to check that their own institution did not have any ethics procedures in place that would over ride those incorporated into the design, and asked if they had any suggestions for modifications. They were also informed that if they wished to they could modify the question lists to add anything to reflect their own unique library design and resource collections and provide further information for their own personal institute investigations. All data were agreed to be returned to Huddersfield for analysis of themes.

Each institution arranged for their own focus groups, sending a pre-designed email inviting students to participate, with an incentive to the value of £10. Most institutes chose to offer print or photocopying credit, but due to the nature of project timing restrictions, to encourage attendance by final year students some offered coffee gift vouchers or similar commercial incentives.

Focus group feedback is still in the process of being returned to Huddersfield, but initial contact indicates that each collaborator has had varying success rates with participation levels, regardless of the nature of the incentive offered. One institution received responses from 209 students interested in participating, while another had only a very small number of students reply.

Non/low Use at the University of Huddersfield

Huddersfield has collected over five years worth of data on library usage. A separate in-house working party has been set up alongside the Library Impact Data Project to progress the non/low use agenda. The aim of this group is to increase attainment levels by engaging non/low users where appropriate and working progressively with staff to embed library use. To this extent academic librarians at Huddersfield are also working closely with tutors on a selected sample of courses to explore the reasons for unexpectedly low use of library resources; the courses were agreed by

the University's Quality Standards Advisory Group (QSAG) in December 2010.

In addition to the focus group work described above, separate focus groups have been held with students on selected 'non/low use' courses. It is important to note that any findings from these focus groups may reflect a worst case scenario from those that may not engage with library resources. Those students' opinions need to be seen in this light and used to advise us on service improvements, not to highlight poor service. However, they may prove a useful comparison with the LIDP focus groups, which by definition will include users who engage.

Short-term objectives are to flesh out themes from the focus groups to advise on areas to work on and to check the amount and type of contact subject teams have had with the specific courses in order to compare library teaching hours to attainment (with the caveat that poor attainment does not reflect negatively on the library support). It is hoped that focus groups' themes can help to identify areas for improvement in order to target promotion and increase tutor awareness. Data for these courses will also be checked for a correlation across different years to allow specific targeting of library information skills sessions, e.g. 1st or 2nd years etc. This will allow the service to target precious staff resources in an appropriate and efficient manner to ensure that support for students is available at key times. Profiling of students and courses will also begin with a review of reading lists to see the amount of use. It is also intended to compare Computing and Library Services (CLS) surveys and the National Student Survey (NSS) with the chosen courses to see if there is a connection.

In September 2011, a baseline questionnaire or exercise for new students will be created to establish the level of their information literacy skills. This will take into account the tendency for Net Generation students to overestimate their own skills and then demonstrate poor critical analysis once they get onto resources (Weiler, 2005). It will also be used to inform use of web 2.0 technologies for different cohorts, e.g. health vs. computing. New or repeat focus groups will be held to check the progress of the project.

It is hoped that by refining the data and targeting of appropriate staff resource at the point of need will help to increase student attainment over time.

Further Research

The project has had to remain focussed because of the six-month timescale. However, through analysis of the data and in discussion with partners and other interested parties, a number of potential avenues of future research have been identified.

Most obvious is the link back to the original non/low use project at Huddersfield as described above. Other data from Huddersfield have highlighted a particular trend as shown in Figure 5. It appears that the average number of books issued over the last 5 years has stayed relatively stable for upper and lower second-class degrees, but has risen substantially for first class degrees and dropped for third class degrees. There would clearly be some merit in exploring this further.

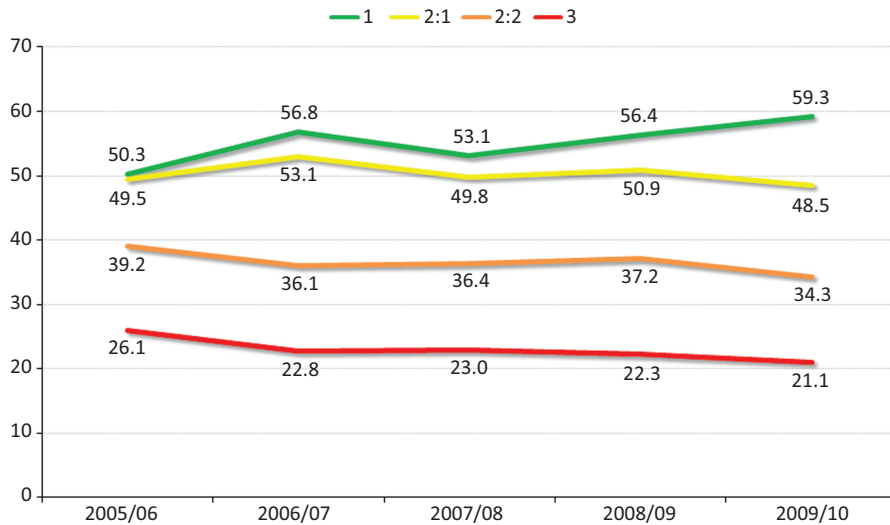
Furthermore, the project has only counted the number of book loans and e-resource access in a given year and not the frequency over time. This would mean a significant amount of work as the data supplied by each partner would need to be re-submitted. However, it may add a further dimension to the study.

Comments from the recent SCONUL conference suggest that further analysis by gender would also be of interest. The project does have one set of data that includes this as well as information on country of origin (British/European/'Overseas') and the team intends to run this as a test sample if time permits at the end of the project.

There has also been some discussion on the value added by libraries and universities as a whole. A future project could use data from the point of entry, e.g. UCAS tariff, and map this against library usage and final award. This could potentially show the value-added benefits, for example, student X who entered university with high grades and left with high grades may show a similar level of usage to student Y who entered with lower grades, but left with a first class degree. Analysis of library usage could give credence to the argument that a significant amount of value had been added by the library for student Y over student X.

Further work is also needed on baseline surveys to measure what level students enter university as this too will have a potential affect on library

Fig. 5: Book issues by final degree result.



usage. Another suggestion from the SCONUL conference was to investigate the socio-economic background of the student to see if this had an impact.

There needs to be more investigation of e-resource usage as the method of measuring via Athens access or similar is a crude but common measure. Further work would need to be undertaken in conjunction with a publisher to track usage at a more granular level, although this would raise further legal issues over data protection. However, it would be beneficial to both publishers and universities to see which journal titles were heavily used by researchers, undergraduates or a mixed audience.

Finally, although the project team have been very clear to state that the correlation between library usage and student attainment is not a cause and effect relationship, it does beg the question that if use of resources does benefit student attainment, what happens if a budget cut reduces those resources significantly? It has been suggested that this project could prove a powerful argument for library directors when trying to negotiate budget.

Conclusion

The project is very hopeful that a correlation for at least some of the usage data supplied can be shown and that the hypothesis that *'there is a statistically significant correlation across a number of universities between library activity data and student attainment'* can be proved.

The next step for the project is to release the data for others to exploit. The aim is to do this using an Open Commons Data Licence. Ideally, the project would like to release each institution's data separately; however, this will need the unanimous approval of the project team and senior management within the partner libraries. All partners are due to be sent a report on their data in the coming weeks and a decision will be reached by mid July. The main concern is that if one partner does not show a statistical significance but the project as a whole does, this may reflect negatively on that institution, despite there not being a cause-and-effect relationship. However, if this is the case, the intention is to release the data as a single set.

The project posts regular items on the project blog and will release a final report towards the end of July. It is intended to keep the blog open after the life of the project in order to post updates and further research.

Acknowledgements

The project has been a model example of how eight universities can, in a relatively short space of time, collaborate on a shared service. This is down to the significant contribution of all project team members: Dave Pattern, Bryony Ramsden, Phil Adams, Leo Appleton, Iain Baird, Polly Dawes, Regina Ferguson, Pia Krogh, Marie Letzgus, Dominic Marsh, Habby Matharoo, Kate Newell, Sarah Robbins, Paul Stainthorp and Andy McGregor at the JISC. Details of all members of the project team can be found on the Library Impact Data project blog.

References

De Jager, K. (2002a): 'Impacts & outcomes: searching for the most elusive indicators of academic library performance'. *Proceedings of the 4th Northumbria International Conference on Performance Measurement in Libraries and Information Services "Meaningful Measures for Emerging Realities"*, pp. 291–297.

- De Jager, K. (2002b): 'Successful students: does the library make a difference?', *Performance Measurement and Metrics*, 3(3), pp. 140–144.
- Dent, V.F. (2006): Observations of school library impact at two rural Ugandan schools', *New Library World* 107(9/10), pp. 403–421.
- Goodall, D. and D. Pattern (2011): 'Academic library non/low use and undergraduate student achievement: a preliminary report of research in progress', *Library Management*, 32(3).
- Gross, M. and D. Latham (2007): 'Attaining information literacy: an investigation of the relationship between skill level, self-estimates of skill, and library anxiety', *Library and Information Science Research*, 29, pp. 332–353.
- OCLC (2006): *College Students' Perceptions of the Libraries and Information Resources: A Report to the OCLC Membership*, OCLC, Dublin, OH.
- Ontario Library Association (2006): *School libraries and student achievement in Ontario*. Ontario Library Association, Ontario [Online] Available at <http://www.peopleforeducation.com/school-libraries/2006> (accessed 21 June 2011).
- Pattern, D. (2010): Non/low library usage and final grades. Self-plagiarism is style blog [Online] Available at: <http://www.daveyp.com/blog/archives/1385> (accessed 21 June 2011).
- Ramsden, Bryony (in press): 'Evaluating the impact of learning space', *Reference Services Review* 39(3).
- Rowlands, I., D. Nicholas, P. Williams, P. Huntington, M. Fieldhouse, B. Gunter, R. Withey, H.R. Jamali, T. Dobrowolski and C. Tenopir (2008): 'The Google generation: the information behaviour of the researcher of the future', *Aslib Proceedings*, 60(4), pp. 290–310.
- Weiler, A. (2005): 'Information-seeking behavior in generation Y students: motivation, critical thinking, and learning theory', *Journal of Academic Librarianship*, 31(1), pp. 46–53.
- White, S. and G. Stone (2010): 'Maximising use of library resources at the University of Huddersfield', *Serials*, 23(2), pp. 83–90.

Websites Referred to in the Text

Lord Browne's Review of Higher Education Funding and Student Finance: <http://hereview.independent.gov.uk/hereview/> (accessed 21 June 2011).

Activity data: JISC: <http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata.aspx> (accessed 21 June 2011).

A Preliminary Report on the Library Impact Data Project

Open Data Commons: <http://www.opendatacommons.org/> (accessed 21 June 2011).

Library Activity Data Project: <http://library.hud.ac.uk/blogs/projects/lidp/> (accessed 21 June 2011).

JISC Legal: <http://www.jisclegal.ac.uk/> (accessed 22 June 2011).

Using OpenURL Activity Data Project: http://edina.ac.uk/projects/Using_OpenURL_Activity_data_summary.html (accessed 21 June 2011).