University of Huddersfield Repository

Stone, Graham, Pattern, David and Ramsden, Bryony

Library Impact Data Project: hit, miss or maybe

Original Citation

Stone, Graham, Pattern, David and Ramsden, Bryony (2012) Library Impact Data Project: hit, miss or maybe. In: Proving value in challenging times: proceedings of the 9th Northumbria international conference on performance measurement in libraries and information services. University of York, York, pp. 385-390. ISBN 978-0-901931-11-5

This version is available at http://eprints.hud.ac.uk/15563/

http://eprints.hud.ac.uk/

The Library Impact Data Project: hit, miss or maybe

Graham Stone, Dave Pattern and Bryony Ramsden

Abstract

Purpose
In February 2011 the University of Huddersfield along with 7 partners were awarded JISC funding through the Activity Data programme to investigate the hypothesis that:

"There is a statistically significant correlation across a number of universities between library activity data and student attainment"

The Library Impact Data Project aimed to analyse users' actions with regards to library usage and then linking those to final degree award. By identifying a positive correlation in this data those subject areas or courses which exhibit high usage of library resources can be used as models of good practice.

Design, methodology or approach
The overall approach of the project is to extract anonymised activity data from partners' systems and analyse the findings. For each student who graduated in the sample years, the following data was required: final grade achieved; number of books borrowed; number of times e-resources were accessed; number of times each student entered the library and school or faculty. This data was then collated, normalised, and then analysed. In addition all partners were asked to hold a number of focus groups in order to secure qualitative data from students on library usage to provide a holistic picture of how students engage with library resources.

Findings
This paper will report on the findings of the project which ran from February to July 2011. It will consider whether the hypothesis was proven for the three indicators of library usage.

Research or practical limitations or implications
The main aim of the project was to support the hypothesis. The project acknowledges however, that the relationship between the two variables is not a causal relationship and there will be other factors which influence student attainment.

Conclusions
The paper will discuss the implications of the results and suggest further work that could result from the projects findings.

Uncontrolled keywords: Library usage, student attainment, low use, non-use, academic libraries, undergraduate students, achievement

Background

In 2009/10, data at the University of Huddersfield was analysed for over 700 courses over four years - 2005/6 – 2008/9. This investigation suggested a strong correlation between library usage and degree results, and also significant under usage of expensive library resources at both School and course level. Three indicators of library usage were used:

- Access to e-resources
- Book loans
- Access to the library

Data was then matched against the student record system (SITS:Vision) and anonymised. The correlation between library usage and final degree was particularly strong in relation to access to e-resources and book loans, and existed at School and course level.

In April 2010 the initial findings were presented at the UKSG Conference (White and Stone, 2010). It was highlighted that the correlation between library usage and grade had not yet been significance tested and that it was not known whether Huddersfield was an anomaly or the norm. To this extent, other academic institutions were invited to benchmark against the findings. As a result, a number of universities from across the UK higher education sector expressed an interest in becoming partners in this project.

Library Impact Data Project

In February 2011 the University of Huddersfield along with 7 partners; University of Bradford; De Montfort University; University of Exeter; University of Lincoln; Liverpool John Moores University; University of Salford and Teesside University successfully bid and were awarded JISC funding through the Activity Data programme (JISC, 2011) to investigate the hypothesis that:

"*There is a statistically significant correlation across a number of universities between library activity data and student attainment*"

The Library Impact Data Project (LIDP) (Various, 2011) aimed to analyse users' actions with regards to library usage and then linking those to final degree award. By identifying a positive correlation in this data those subject areas or courses which exhibit high usage of library resources can be used as models of good practice.

Data Requirements

By partnering with universities that represent a cross-section of size and mission the project aimed to provide a rich data set on which to work. Table 1 shows the list of requirements the project asked of its collaborators. The minimum data requirements for participation in the project are highlighted in bold.

Academic year of graduation	e.g. 2009/10
Course title	Software Development
Length of course in years	3
Type of course	post grad
Grade achieved	2:1
School/academic department	School of Computing & Maths
Number of items borrowed from library - either the total number borrowed by that student - or separate values for each	e.g. 50 items during the 3 years of the course 11 items in 2007/8, 16 in 2008/9, 23 in 2009/10

academic year	
Number of visits to the library • 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) • either the total number of logins made by that student • or separate values for each academic year	

Table 1: Data requirements

In addition, some of the partners submitted extra data such as network logins. As the research centred on undergraduate students, partners were requested either to extract their data independently of other students or to provide a key so that the data could be extracted at Huddersfield. Year of graduation was only necessary if partners intended to produce longitudinal analysis across different intakes of students.

The data extract was required in an Excel-readable format, to aid preparation and ease of transfer to SPSS/PASW for analysis.

Legal Issues

From the outset, a major issue for the project was to ensure it was abiding to legal regulations and restrictions. At the initial bidding stage, partners were asked to discuss privacy issues with their institution's legal advisor, records manager and/or ethics committee.

In order for Huddersfield to handle the data, the UK Data Protection Act required that the data provided should not include any information that could be used to identify any named individual, e.g. their name or a campus network ID. In addition, any small courses were excluded to prevent identification of individuals i.e. where a course had less than 35 students and/or fewer than 5 of a specific degree level.

The project contacted JISC Legal prior to data collection to confirm procedures were appropriate, and additionally liaised with the University of Huddersfield Records Manager and the University's legal advisor. Advice was also taken from another project in the strand, the UK OpenURL Router Services (Edina, 2011), which uses the following statement to notify library and resources users of 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."

Focus groups were also provided with a briefing and a consent form to ensure participants were fully aware of how the anonymised data from the group was to be used and advising them that they could leave the group at any point (see below).

Results

Due to limitations on what data format was available, testing methods took some time to refine. While usage data was continuous, the final mark was not, and took the form of degree format rather than a percentage score. As a result, following several attempts using various methods, the Kruskal-Wallis (KW) test was selected, in combination with the Mann Whitney U test (MW). These tests combined analyse differences between groups of data, the former checking for differences between groups overall without specifying where differences lie, the latter allowing for several tests to be conducted analysing differences between specific groups. The size of the difference can then be measured using a simple manual calculation.

The tests themselves are not without problems, and follow a convoluted process. Before the KW test can be run, the data must be tested to ensure it does not follow a normal, bell curved distribution. Following the KW test, data must be analysed visually using boxplots in order to decide what degree results to compare in the MW tests (see figure 1).
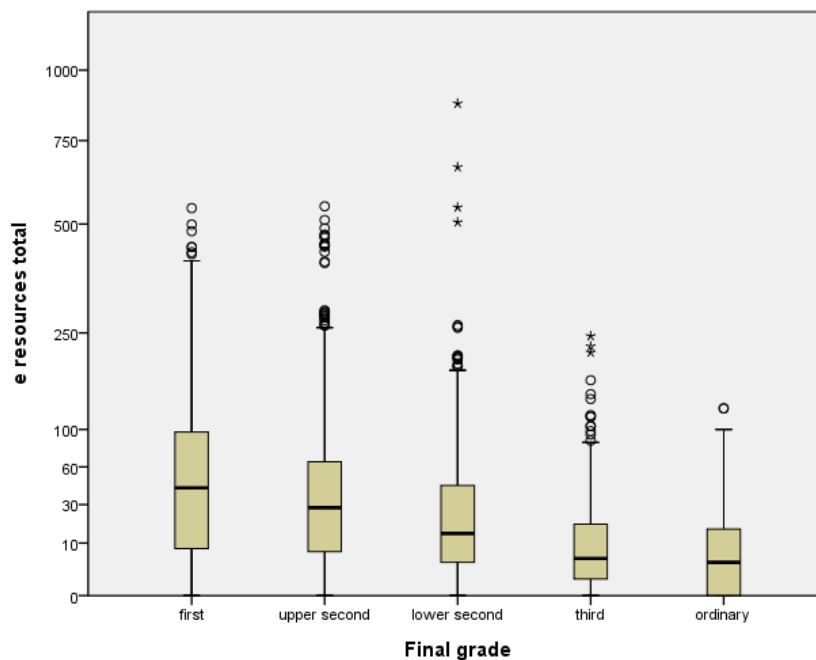

Figure 1: A boxplot of e-resource usage, indicating a difference in usage between high and low degree results.

However, the number of MW tests that can be conducted is restricted as the more the test is run on one set of data groups, the more likely a difference will be found accidentally, and the higher the significance level is required. Had the data set been smaller (for example of a specific course), analysis could have been conducted to indicate whether there was a specific direction of usage and thus a decrease in degree result as usage declines. However, as the combined data for all institutions measured in excess of 33,000 records, and analysis for individual institutions was conducted across all undergraduate degrees, it was not possible to measure a directional relationship.

Data analysis indicated that there were differences between degree results in terms of using electronic resources and borrowing items from the library. Effect size, when at a significant

level, indicated that the higher degree had higher usage than the lower degree, albeit to a varying extent. Course level analysis is expected to indicate larger effect sizes, but would naturally be dependent on the course itself: course usage will depend on the nature of the subject of study and the nature of the assignment requirements. The data overall is expected to be skewed somewhat due to the inclusion of 'anomaly' courses such as computing or art, where usage of library resources may not necessarily be common or measurable.
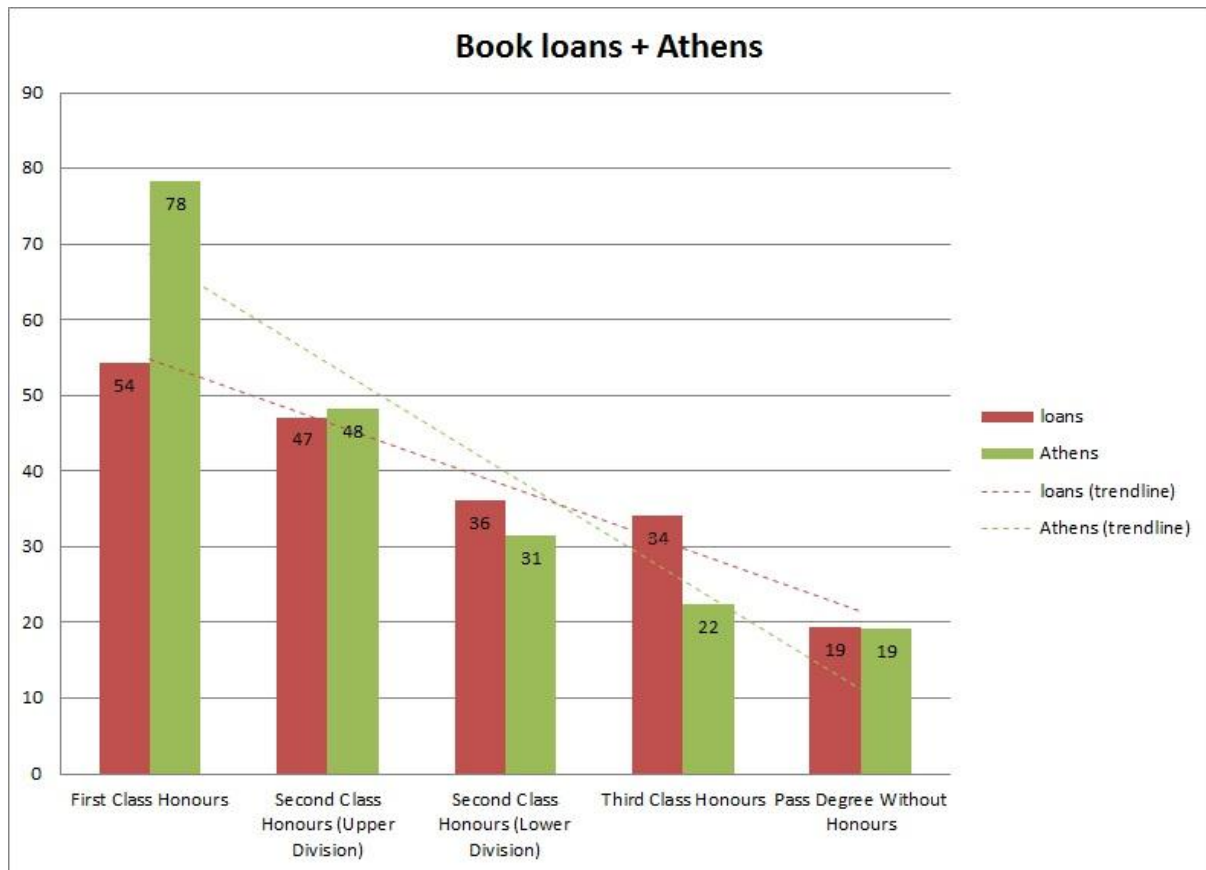

Figure 2: Example of book loans and Athens logins vs. degree classification

Figure 2 shows a typical result from one of the project partners based on averages. Our research supports the visual representation of variance between degree results and usage, and does so across a range of data e.g. subjects, which supports our hypothesis that library usage does impact on students' attainment.

Focus groups
Focus groups were conducted to seek out qualitative data as to why some courses or students would or would not use library resources, and were considered particularly useful for investigating anomaly courses further. All institutes were asked to conduct groups, and were provided with:

- guidelines for how to conduct the group
- a script to introduce the purpose of the group
- a statement of research purpose with contact details and clearly indicating students could leave at any point
- consent forms

- questions modifiable by the host institution to reflect their resources and/or add any other questions they chose
- a short survey of library resource usage to compare to the responses provided during the discussion

Groups were advertised to students in various ways, but all included an incentive to attend in either print/photocopying credits or vouchers, both to the value of £10. Different institutions had varying success in gathering participants, but most were able to hold at least one small group. All institutions were asked to ensure their own ethical guidelines were fully adhered to, but were provided with materials based on the guidelines at Huddersfield.

Questions for the groups covered a wide variety of subjects regarding frequency of resource use, how they felt their usage levels compared to others on their course, whether they had ever encountered any difficulties using resources, and what they felt was required for a good degree result. Additionally they were asked about previous experience of library usage, be it of public libraries or of an educational nature, and how much reading they conducted outside of that required by their lecturers.

The coding involved a process of making notes of potential themes arising: the process creates a sometimes lengthy list of broad themes with satellite elements of more descriptive themes which can then be designated a code word, with codes refined and reduced down to a manageable number. The code or multiple codes were attached where appropriate within the transcript. Codes were based around themes matching the questions raised, but also included elements relating to the nature of using library facilities as a study environment. Due to time limitations on the project, coding analysis was limited to studying the frequency of an item being discussed. While this is a very simplistic method of analysing the data, it does emphasise just how important a library resource or feature is to the student, regardless of whether their comment is positive or negative: if a student mentions they have had difficulties using a resource several times during the group, it suggests that their difficulties have made an impact on many levels, be it of their perception of the resource, of library provisions, or of support.

Results varied according to the nature of the institution and the number of participants, but as a whole, students frequently discussed their information seeking process whether it involved library provided resources or not. They often had proactive, systematic techniques, yet varied in confidence of their skills in finding information. As a whole, library resources were often a point of in-depth discussion indicating that they were important to students, regardless of whether they perceived difficulties in using them or not. Technology was unsurprisingly integral to this information seeking process, with students discussing specific software packages they used, as well as visiting the library to use the computing facilities in general.

Toolkit
One of the planned outputs of the LIDP was the release of the types of activity data required and methodology required for others to test their data. This took the form of a toolkit which was released in September 2011 (Stone, 2011a). The toolkit describes how other universities can capture their data in order to compare results as described above. It also discusses the data protection and legal issues which may be encountered.

A significant part of the toolkit also gives advice on the types of statistical testing required in order to prepare and analyse the data. Appendices are included to assist universities in running the various focus groups discussed above; these include a consent form, a list of questions used by the project partners and a quick survey.

In order to keep within the spirit of the project, which undertook to release the data collected, the toolkit provides guidelines for the release of data under an Open Data Commons Attribution License (Open Knowledge Foundation, 2011). Universities that choose to release the data in this way have been offered a link on the LIDP blog for others to use.

Further work

Non-low use

The Library Impact Data Project had a finite goal to investigate the hypothesis. However, as the project developed, a number of possible further studies were suggested, not least that by supporting the hypothesis for book loans and e-resource usage across all the partners the project could return to the inspiration for LIDP, the non/low use project at Huddersfield (Goodall and Pattern, 2011). Now that the project has shown that there is a statistical significance between usage and attainment there is an opportunity to engage non/low users on selected courses in order to understand why students do not engage.

One way of taking this idea further would be to use data already available to see if there is a relationship across all years of the courses. The initial non/low use study uncovered some interesting data usage by course year. Some courses have no statistical significance with final grade in year one, but others do. By delving deeper into this, staffing resources could be targeted more effectively to help students at the point of demand. To support this it would also be necessary to check the amount and type of contact subject teams have with the specific courses to compare library contact hours to attainment, although this would not reflect a measure of cause and effect for library staff!

Furthermore, there is an opportunity to investigate the use of library reading lists at Huddersfield (Pattern, 2011), for example, does wider reading lead to better attainment? Originally the project wanted to look into results from the National Student Survey (NSS), although this was not possible due to data protection issues for all the partners, however, individual universities could compare the NSS data to courses where there is high non/low use.

Business Intelligence

Numerical data may also be utilised for comparing socio-economic and demographic information at course, school and university wide levels to investigate whether there are any effects of gender, nationality (UK, other European and international could certainly be investigated); distance learning vs. on campus learning; sandwich and placement extended undergraduate courses vs. standard 3 year undergraduate courses. Qualitative data could also be utilized to modify and improve services, reflect on student support facilities and networks, and collection management considerations.

This was a popular theme in questions at the SCONUL and LIBER conferences. Some of these ideas have also been discussed at the recent Business Librarians Association Conference (Stone, 2011b). In addition, business intelligence could be used to help provide data for library directors, e.g. for the scenario 'if budget cuts result in less resources, does attainment fall'? In fact the Library Impact Data Project was recently cited in SCONUL's recent response to the Higher Education White Paper "Higher Education: Students at the Heart of the System" (SCONUL, 2011).

Conclusion

As work on the project has progressed, a number of similar projects have also been undertaken around the world, such as research undertaken by Jantti and Cox (2011) at the University of Wollongong in Australia and the recent report form the Association of College

and Research Libraries (2010) in the United States. Early results show similar results to the LIDP, which adds to the evidence and gives further credence to support the hypothesis that:

"*There is a statistically significant correlation across a number of universities between library activity data and student attainment*"

Word count: 2,748

References
Association of College and Research Libraries (2010) *Value of Academic Libraries: A Comprehensive Research Review and Report*. Researched by Megan Oakleaf. Chicago: Association of College and Research Libraries.

Edina (2011) About the *OpenURL Router* [online] Available at: http://openurl.ac.uk/doc/ [accessed 31 October 2011]

Goodall, D. and Pattern, D. (2011) 'Academic library non/low use and undergraduate student achievement: a preliminary report of research in progress'. *Library Management*, 32 (3). pp.159-170. Available at: http://dx.doi.org/10.1108/01435121111112871, http://eprints.hud.ac.uk/7940/ [accessed 31 October 2011]

Jantti , M. and Cox, B. (2011) 'Capturing Business Intelligence required for targeted marketing, demonstrating value, and driving process improvements', In: *9th Northumbria International Conference on Performance Measurement in Libraries and Information Services: Proving value in challenging times*, 22-25 August 2011. University of York.

JISC (2011) *Activity Data* [online] Available at: http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata.aspx [accessed 31 October 2011]

Open Knowledge Foundation (2011) *Open Data Commons Attribution License (ODC-By) v1.0* [online] Available at: http://opendatacommons.org/licenses/by/1.0/ [accessed 31 October 2011]

Pattern, D. (2011) *MyReading project* [online] Available at: http://library.hud.ac.uk/blogs/projects/myreading/ [accessed 31 October 2011]

SCONUL (2011) *Society of College, National and University Libraries (SCONUL) response to the Higher Education White Paper "Higher Education: Students at the Heart of the System"*, 19 September 2011. Available at http://www.sconul.ac.uk/news/he_whitepaper/hewhitepaper_response.doc [accessed 31 October 2011]

Stone, G. (2011a) *Library Impact Data Project toolkit* [online] Available at: http://library.hud.ac.uk/blogs/projects/lidp/2011/09/28/library-impact-data-toolkit/ [accessed 31 October 2011]

Stone, G. (2011b) *Talking to Business Librarians at the BLA Conference* [online] Available at: http://library.hud.ac.uk/blogs/projects/lidp/2011/07/15/talking-to-business-librarians-at-the-bla-conference/ [accessed 31 October 2011]

Various (2011) *Library Impact Data Project* [online] Available at: http://library.hud.ac.uk/blogs/projects/lidp/ [accessed 31 October 2011]

White, S. and Stone, G. (2010) 'Maximising use of library resources at the University of Huddersfield'. *Serials.* 23 (2), pp.83-90 Available at: http://dx.doi.org/10.1629/2383, http://eprints.hud.ac.uk/7811/ [accessed 31 October 2011]

Author contact details

Graham Stone
Information Resources manager
University of Huddersfield
g.stone@hud.ac.uk

Dave Pattern
Library Systems Manager
University of Huddersfield
d.c.pattern@hud.ac.uk

Bryony Ramsden
Research Assistant
University of Huddersfield
b.j.ramsden@hud.ac.uk