



University of HUDDERSFIELD

University of Huddersfield Repository

Stone, Graham, Ramsden, Bryony and Pattern, David

Library Impact Data Project Toolkit

Original Citation

Stone, Graham, Ramsden, Bryony and Pattern, David (2011) Library Impact Data Project Toolkit. Manual. University of Huddersfield, Huddersfield. (Unpublished)

This version is available at <http://eprints.hud.ac.uk/id/eprint/11571/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

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



Library Impact Data Project Toolkit



This work is licensed under a
Creative Commons Attribution 3.0
Unported License

Library Impact Data Project Toolkit

The Library Impact Data Project, which is part of the JISC Activity Data programme sought to prove the hypothesis that:

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

The project ran from February – May 2011 and involved the following institutions project:

- University of Bradford
- De Montfort University
- University of Exeter
- University of Huddersfield
- University of Lincoln
- Liverpool John Moores University
- University of Salford
- Teesside University

The aim of this project is to prove a statistically significant correlation between library usage and student attainment. By identifying subject areas or courses which exhibit low usage of library resources, service improvements can be targeted. Those subject areas or courses which exhibit high usage of library resources can be used as models of good practice.

Further discussion on the projects results and a complete set of anonymised data can be found on the project blog: <http://library.hud.ac.uk/blogs/projects/lidp/>

One of the outcomes of the project was to provide a toolkit to assist other institutions who may want to test their own data against this hypothesis. The toolkit aims to give general guidelines about:

1. Data Requirements
2. Legal Issues
3. Analysis of the Data
4. Focus Groups
5. Suggestions for Further Analysis
6. Release of the Data

LIDP data has been made available under the Open Data Commons Attribution License
<http://opendatacommons.org/licenses/by/1.0/>

If you have used this toolkit to look at your data, we ask you to share your data too. Please let us know and we will link to it from the project blog.

1. Data Requirements

The following is a list of requirements we made of our collaborators. Requirements will vary according to your own institution and the depth of analysis you wish to conduct. The minimum requirements for your data following the processes detailed here would be the items in bold.

For a specific academic year (e.g. 2009/10), extract details for each graduating student i.e.

academic year of graduation	e.g. <i>2009/10</i>
course title	<i>Software Development</i>
length of course in years	<i>3</i>
type of course	<i>post grad</i>
grade achieved	<i>2:1</i>
school/academic department	<i>School of Computing & Maths</i>
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	e.g. <i>50 items during the 3 years of the course</i> <i>11 items in 2007/8, 16 in 2008/9 and 23 in 2009/10</i>
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	

You may also wish to consider analysing other data not listed here, such as network logins. LIDP research centred around undergraduate students, so if it is possible to extract their data independently of other students, course type would not be needed. Similarly, the length of course would not be required if all honours courses can be extracted automatically. Year of graduation would only be necessary if you intend to produce longitudinal analysis across different intakes of students.

Notes

- As per the UK Data Protection Act, the extract shouldn't include information that identifies, or could be used to identify, a named individual (e.g. their name or a campus network ID).
- Ideally, the data extract should be in an Excel-readable format, to aid preparation and ease of transfer to SPSS/PASW for analysis.

Once data has been successfully extracted, it will need to be prepared for use in the analytical software (see section on SPSS/PASW).

2. Legal Issues

One of the big issues for the project was to ensure we were abiding to legal regulations and restrictions, and continue to do so. You should ensure you discuss privacy issues with your institute's legal advisor, records manager and/or ethics committee. As detailed earlier we made efforts to ensure there is:

- full anonymisation of both students and universities so that neither can be identified via the data. We contacted JISC Legal prior to data collection to confirm our procedures were appropriate, and additionally liaised with our Records Manager and the University's legal advisor.
- **We have excluded any small courses in public reports or open access release to prevent identification of individuals i.e. where a course has less than 35 students and/or fewer than 5 of a specific degree level.**
- To notify library and resource users of our data collection. We referred to another data project, EDINA, which provides the following statement for collaborators to use on their webpages:

“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.”

http://edina.ac.uk/projects/docs/Appendix_D_How_to_inform_your_users_about_data_processing.pdf

Focus groups have also been conducted with a briefing and a consent form to ensure participants are fully aware of data use from the group and of their anonymisation and advising them that they can leave the group at any point (see section on focus group guidelines).

3. SPSS/PASW Analysis of Data

Preparation of data

Data should be readied for use in the software by modifying it slightly. Degree results should be changed to represent a single digit number e.g. a first class degree is 1, an upper second class degree is 2, lower second class 3, and so on, so that each student can be easily identified as belonging to a specific group of results in the software. SPSS/PASW can only deal with the degree data if it follows a numerical sequence, rather than based on text: it will reject any text it finds, so any listing of fail or ordinary degree result should also be modified to be represented numerically.

Analysis process

Choice of test is very much dependent on what form the data takes. For the tests used on LIDP data, there would need to be:

- The final degree result i.e. honorary and ordinary degrees. Passes and fails and any other non-honorary system would potentially be problematic for this technique, as there wouldn't be enough variation in results. Ideally it would include 3 or more variants of results for grouping data.
- Usage information in numerical format.

Analysis is based on specifically locating differences between groups of students (i.e. by degree result) in their resource usage. The analysis process incorporates several tests, used either to prove the data follows a specific pattern and thus the appropriateness of later tests, or to measure data for differences. All applicable tests are measured for significance at a 5% value ($p < 0.05$) unless otherwise stated.

1. **Kolmogorov-Smirnov test**: used to prove the data is not normally distributed. Normal distribution follows the famous bell curve shape which when charted shows a peak in the centre. If data is normally distributed, an alternative test is required for analysis.
2. **Kruskal-Wallis (KW) test**: indicates that there are differences between groups (i.e. degree results). The test will not, however, indicate where differences lie. Additionally, it will indicate a difference when only one group shows different results to the others.

Ranks

	Degree result	N	Mean Rank
e resources total	First class	438	2262.26
	Upper second	1472	2045.53
	Lower second	1250	1727.16
	Third class	316	1288.74
	Ordinary	207	1043.44
	Total	3683	

Test Statistics^{b,c}

		e resources total
Chi-square		339.637
df		4
Asymp. Sig.		.000
Monte Carlo Sig.	Sig.	.000 ^a
	99% Confidence Interval	
	Lower Bound	.000
	Upper Bound	.000

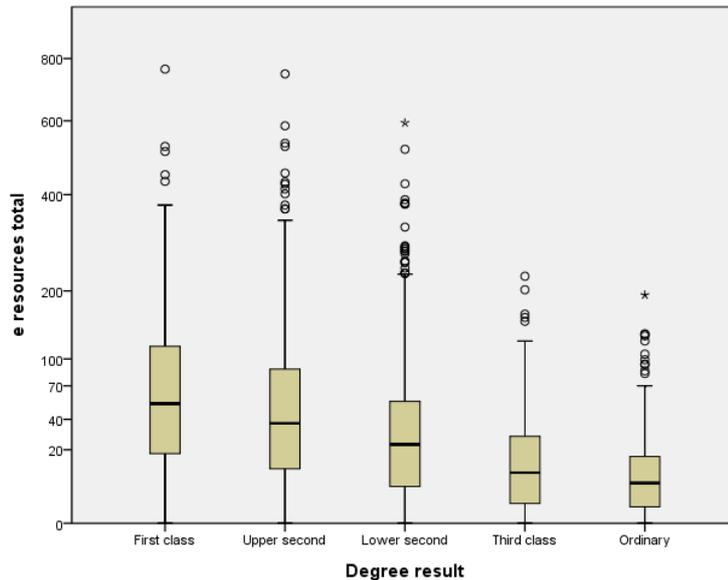
a. Based on 10000 sampled tables with starting seed 1487459085.

b. Kruskal Wallis Test

c. Grouping Variable: Degree result

Both the test's resulting chi-square value and significance levels must be at an appropriate level to prove that the test indicates differences between degree results in the data being analysed. While we can make a fair conclusion about whether there is a difference between usage using the mean ranks (see the ranks in the table above), it must be proven to be a statistically significant difference. The tables above indicate differences between e-resource usage in each degree result, with a very large chi-square value much higher than the required critical value for a significant result (a sample chi-square value table can be found at <http://people.richland.edu/james/lecture/m170/tbl-chi.html>). The significance level is very close to zero, and is supported by the Monte Carlo significance (see below for further information on the Monte Carlo estimate), indicating minimal likelihood of the critical value occurring by chance.

3. **Graphical representation:** a boxplot is used to assess which degree result comparisons are appropriate or of interest. The boxplot will indicate if the KW test is showing a difference because of one group. Assessment for selecting differences/similarities to analyse should be based primarily on the central 50% (i.e. the box itself) and the median value (the dark black line within the box).



4. **Mann-Whitney (MW) test:** based on choices from the boxplot, this test is conducted on a limited number of comparisons to test whether there are differences between specific degree results. The more comparisons are conducted, the stronger the significance required to validate the results i.e. 3 tests based originally on 5% significance would require a significance of 0.167 or less to be valid.
5. **The effect size:** a simple manual calculation to measure the size of the effect based on the Mann-Whitney data. The effect is calculated by dividing the Z value from the Mann-Whitney test by the square root of the sample size used in that MW. Effect sizes are usually measured as per Cohen's values:
- 0.1-0.23 = small
 - 0.24-0.36 = medium
 - 0.37 or more = large.

A small difference, it should be noted, should not be interpreted as a negative result, as it is still a statistically significant difference.

NB: for very large or small samples, it is recommended a [Monte Carlo estimate](#) is used in the KW and MW tests. The estimate creates a simulated sample matching the distribution of your own data and repeatedly tests it for significance, thus ensuring the significance of the tests is correct.

You may also wish to consider using the Jonckheere-Terpstra test in future analyses to measure whether there is an ordered pattern throughout the degree results i.e. a decrease in usage as degree results drop. The test performs best with smaller numbers, and so could potentially work well on course or school level.

We referred to the following text for guiding our use of SPSS:

Field, Andy P. (2009) *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)*. 3rd Ed. London: Sage.

4. Focus groups

Guidelines

For the purposes of LIDP, questions were designed to examine:

- How much students used the library, and whether they perceived that as different to other students they knew
- How and where they used resources, and whether they experienced any difficulties doing so
- To what extent the library fulfilled their information and learning space requirements.

The nature of these themes can be modified and added to according to resources and facilities provided, or to reflect specialised requirements at course level.

Students were asked to volunteer via email, advertising a small reimbursement for their time, which may include shopping gift vouchers or print/photocopying credits. Success in response rate and attendance will vary according to the term schedule and the size of the reimbursement, so it is advised to plan group sizes accordingly: assume that 50% of those asked will attend the group but make plans for 100%! The length of the session should be detailed in the email, and scheduled so that participants can easily fit it into their timetables and leave early if required. An hour with forthcoming students can create a substantial amount of data (on average around 15-20 sides of transcription).

Focus groups should be conducted following any ethical requirements of your institution, but should be sympathetic to the anonymity of participants throughout the recruiting, recording, transcribing and data reporting processes. Ensure attendees are clearly informed of the purpose of the group and data collection, and give them the option to leave the group and have their answers deleted at any point during, or following the group meeting. Attendees should also be provided with an information sheet detailing again the purpose of the group, with contact information for those leading the research and for at least one other member of staff so that any complaints or sensitive issues can be raised freely outside of the focus group. See Appendices 1-4.

Coding focus group transcripts

The coding process will depend entirely on the nature of the discussion within each group, and it is advisable to create codes and theme classification during several readings of the full set of data gathered. LIDP used 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. The code, or multiple codes are attached where appropriate within the transcript, for example the following statement covers a large number of themes:

Student: I like to use the library for the Macs in the silent area. I use the design software, but I like how they are near the interior design books as it makes it easy to find stuff I need if I suddenly realise I'm missing something.

The statement can be tagged with codes relating to technology use, but also to accessing library books, ease of use and/or proximity requirements. If a student repeatedly discusses a particular code several times in separate statements, it may be counted for each of those times. If time allows, further readings can be made to elaborate on the coding e.g. to draw out underlying implications from the statement (in the statement above, codes could be used to elaborate on speed of access and convenience, as well as questioning whether the silent area is chosen for proximity alone or also related to silent studying).

Once statements are coded, they can be totalled up to discover indicators of importance to the participants, and analysed to find connections. Frequent combinations, such as those of electronic resource use and technical problems or staff support could, while not necessarily directly discussed, indicate underlying issues, in this case computing issues, a gap in information literacy teaching, or even problems with subscriptions. Issues arising in this way could be confirmed with participants if they agreed to being contacted again, or create a point of further research and discussion at a later date.

5. Suggested further uses for using the data collected

Throughout the project we have been considering what we would have liked to have done that fell outside of the remit of the project. We have listed a number of ideas below that could be considered in your use of the data:

- Gender and socio-economic background? Are there any effects of gender, nationality (UK, other European and international could certainly be investigated) and socio-economic background in use and attainment – this could be built into the data collection.
- What types of data are needed by library directors, e.g. for the scenario ‘if budget cuts result in less resources, does attainment fall’? The [Balanced Scorecard](#) approach could be used for this?
- Does the library add value through better use of resources, e.g. :
 - Does a student who comes in with high grades leave with high grades? If so why? What do they use that makes them so successful?
 - What if a student comes in with lower grades but achieves a higher grade on graduation after using library resources? What did they do to show this improvement?
 - Quite often students who look to be heading for a 2nd drop to a 3rd in the final part of their course, why is this so?
 - What about high achievers that don’t use our resources? What are they doing in order to be successful and should we be adopting what they do in our resources/literacy skills sessions?
- Is there a relationship between the data and VLE use

There are also further suggestions for using the data to target non/low users:

- Find out what students on selected ‘non-low use’ courses think to understand why students do not engage
- Check the amount and type of contact subject teams have had with the specific courses to compare library hours to attainment (poor attainment does not reflect negatively on the library support!)
- Use data already available to see if there is correlation across all years of the courses.
- Begin profiling by looking at reading lists
- Flesh out themes from the focus groups to identify areas for improvement
- Look for a connection between selected courses and internal survey results/NSS results
- Create a baseline questionnaire or exercise for new students to establish level of info literacy skills
- Set up new longitudinal focus groups or re-interview groups from last year to check progress of project
- Use data collected to make informed decisions on stock relocation and use of space
- Refine data collected and impact of targeted help

6. Release of the Data

These notes were released to support the release of the project data under the Open Data Commons Attribution License <http://opendatacommons.org/licenses/by/1.0/>

If you have used this toolkit to look at your data, we ask you to share your data too, as such you may want to follow these guidelines. Please let us know and we will link to it from the project blog.

Background

Each of the 8 project partners provided a set of data, based on the initial data requirements document. Not all partners were able to provide data for e-resource logins and library visits, but all were able to provide library loans data.

In order to ensure anonymity:

1. the 8 partners are not named in the data release, instead they have been allocated a randomly selected name (from LIB1 to LIB8)
2. the names of schools and/or departments at each institution have been replaced with a randomly generated ID
3. the year of graduation has been removed from the data
4. where a course had less than 30 students, the course name has been replaced with a randomly generated ID
5. some course names have been “generalised” in order to remove elements that may identify the institution

Grades

The awarded degree has been mapped to the following code:

- A = first (1)
- B = upper second (2:1)
- C = lower second (2:2)
- D = third (3)
- E = pass without honours

Library Usage

Where supplied by the project partner, the following library usage data measures are included:

- ISSUES = total number of items borrowed from the library by that student (n.b. this may include renewals)
- ERES = a measure of e-resource/database usage, e.g. total number of logins to MetaLib or Athens by that student
- VISITS = total number of times that student visited the library

Other Notes

- each graduate has been allocated an randomly generated unique ID
- where the course/school/department name was not supplied, it has been replaced with N/A
- where the measure of library usage was not supplied by the partner, the value is blank/empty

Appendix 1

JISC Library Impact Data Project

Focus Group Consent Form

I have been fully informed of the nature and aims of this research and consent to taking part in it.

I understand that I have the right to withdraw from the research at any time without giving any reason, and a right to withdraw my data if I wish.

I understand that the focus group will be recorded for transcription purposes.

I give permission to be quoted (by use of pseudonym).

I understand that the tape will be kept in secure conditions at the University of Huddersfield.

I understand that my identity will be protected by the use of pseudonym in the research report and that no information that could lead to my being identified will be included in any report or publication resulting from this research.

I understand that someone other than the researcher may transcribe my interview and that data will be analysed at a central location at the University of Huddersfield.

I understand that I can request a copy of the interview transcript and that the researcher will check that I still consent to data being used from the transcript.

Name of participant

.....

Signature

.....

Date

.....

Name of researcher

.....

Signature

.....

Date

.....

Two copies of this consent form should be completed: One copy to be retained by the participant and one copy to be retained by the researcher

Appendix 2

Focus Group Participant Information Sheet

1) Research project title and purpose

The project is called the Library Impact Data Project (LIDP), and is funded by the Joint Information Systems Committee (JISC) as part of their Activity Data strand of projects. More information about JISC and the Activity Data projects can be found here:

<http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata.aspx>

LIDP aims to find a correlation between library resource usage and final degree results, with research being conducted across 8 different universities. If we prove a correlation, the results could make a big difference for the quality of services and facilities we offer students, and we can focus on why some courses have very high or very low usage of the library facilities and work to improve our provisions accordingly.

2) Why have I been asked to take part?

You are part of a representative sample across the university's undergraduate and postgraduate students studying at the participating institutions involved in the project.

3) Do I have to take part?

No. If you decide to take part and then change your mind, you can leave at any point.

4) What do I have to do?

You will be asked to answer a few questions about how you do or don't use library facilities and resources, such as databases and books, and then complete a quick, short survey. The focus group may take up to a maximum of 90 minutes and the survey will take approximately 5 minutes.

Please try to be as honest as possible. The research aims to find out why you chose to use or ignore various library provisions, and is not meant to be critical of your personal use in any way.

5) What are the possible disadvantages/benefits of taking part?

The only disadvantage should be sacrificing your time when you participate. The researchers will compensate you for your time by providing refreshments and/or printing credit (or similar).

6) What if something goes wrong?

If you wish to make a complaint about anything you experience throughout your involvement in the project or following results dissemination, contact the project manager If you feel your complaint has not been handled appropriately you can contact ...

7) Will my data be kept confidential?

Yes – any information you provide will be kept as confidential, and anonymised so that you cannot be identified in any reports or publications. You may find you are provided with a sticker to wear at the group which identifies you as 'student A': the researcher will use this to connect the comments to each individual without jeopardising anonymity.

8) Why do you need my participation?

We would like your opinions and information on how/why you use the library to help us find out why particular courses or groups of people do or don't use the library and its resources. Any patterns we find will help us to revise services, target poor usage courses and check our subscriptions to journals and databases to ensure we provide what you need for your studying. The project may also help other universities not involved project modify and adapt their own services.

9) What will happen to the results of the research?

The results will be disseminated via the project blog:

<http://library.hud.ac.uk/blogs/projects/lidp/>

You will not be identifiable via the results published.

10) Will I be recorded? How will the recording be used?

You may be recorded for the purposes of keeping track of the group discussion, as notes made by the researchers may not be sufficient data alone: the recording will be transcribed for analysis so that any comments you make can be used within appropriate context and will not misrepresent your library use, but you will not be identified by name (you may be referred to as, for example, student A). The recordings will not be used outside of this process without your written consent, and will not be made available to anyone outside of the project.

11) Contacts for further information

If you require further information, email ...

Thank you for your time and participation in this research.

Appendix 3

Questions for focus groups

Preamble

Thank you for attending this focus group. The groups are designed to gather data for a research project funded by JISC (<http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata.aspx>) across 8 different universities on the impact of library usage. We think library usage and final degree results may be linked. If a correlation is proven, the project could make a big difference to the part the library plays in providing a better student experience. The feedback we receive from you will contribute to us assessing how well the library provides for students throughout their course as well as why use of library facilities and resources might be high for some courses and not others. Findings will be disseminated via the project blog (the link is on the information sheet you have been provided).

You have been chosen as part of a representative sample of students across the university. Participation is voluntary: if you decide to take part you should sign a consent sheet and will be provided with an information sheet to take away with you, with contact details should you wish to complain about the process. If you decide to withdraw you can leave at any time and do not have to have or provide a reason. If you participate you will be provided with [print credit/other motivation] as compensation for your time. All information you provide us will be regarded as confidential and will be fully anonymised so that you cannot be identified in any reports or publications.

1. What resources do you use?

[use the following as prompts]

- a. Books?
 - b. E-books?
 - c. Journals?
 - d. Electronic journals?
 - e. Databases?
 - f. Library staff?
2. How often do you use the library? [prompt – do you feel you use the library a lot or a little bit. Perhaps discuss in comparison to their friends to help them gauge their own usage?]
3. When you use e-resources, where do you use them? [e.g. at home, in classrooms, in the library etc. Prompt for why]
4. Where do you use library-provided resources overall? [e.g. at home, in classrooms, in the library etc. Prompt for why]
5. Do you ever have any difficulties using any resources?
6. Have you ever attended any library resource training sessions? Were they useful? [prompt for why if necessary]
7. Do you do any reading outside of recommended reading? If so, how do you find the information?
8. Did you use library facilities where you last studied? [prompts - were they very different from university? Were you provided with materials/asked to purchase items or given a reading list?] – **this question is to try and help us work out how experienced at searching for information the students were on arrival at university**
9. Does working in a library help you work or make it more difficult for you? Why?

These questions will be followed by a short quick survey

Appendix 4

Short focus group survey

1) How often do you visit the library?

Daily 2-3 times a week Weekly
Monthly Termly Never

2) What do you do when you visit?

Borrow books Writing/typing up assignments Researching
Preparation for tutorials Internet (studying) Group work
Email Internet (fun)/Facebook Meeting friends

3) How many books do you borrow a month?

Never borrowed 1-4 5-9 10-14 15-20 20+

4) How often do you access e-resources a month (i.e. electronic journals, databases, Summon [modify this for your own institution]?)

Daily 2-3 times a week Weekly
Monthly Termly Never

5) Have you bought any books for your course? How many?

Never bought 1-4 5-9 10-14 15-20 20+

6) What kind of library space do you like best?

Silent study Group study Discussion rooms Individual desks

7) Are you:

Undergraduate fulltime Undergraduate part time

Postgraduate fulltime Postgraduate part time

8) Anything else to add about your library services?

Thank you for your time!