Investment in e-journals, use and research outcomes

David Nicholas
CIBER Research Limited, UK

Ian Rowlands
University of Leicester, UK

Library Return on Investment seminar
Universite de Lyon , 20-21 February 2014

Investment in e-journals, use and research outcomes

Research questions

Research questions

How have researchers responded to the unprecedented levels and convenience of access to scholarly journals?

Has enhanced access to the literature led to greater productivity, research quality and other outcomes?

Investment in e-journals, use and research outcomes

Research design

top down

database of UK university indicators

bottom up

deep log analysis of UK university departments

Case study institutions

University of Aberdeen

Bangor University

University of Cambridge

University of Edinburgh

University of Manchester

Rothamsted Research

University of Strathclyde

University of Wales Swansea

University College London

Case study subjects

Chemistry and chemical engineering

Earth and environmental sciences

Economics and econometrics

History

Life sciences and agriculture

Physics

Investment in e-journals, use and research outcomes

Massive consumer acceptance and growth

UK universities have taken full advantage of the enhanced provision of e-journals over the past five years.

The graph opposite shows the number of full text article downloads (from all publishers). Downloads are indexed to 100 for the academic year 2003/04 for ease of comparison.

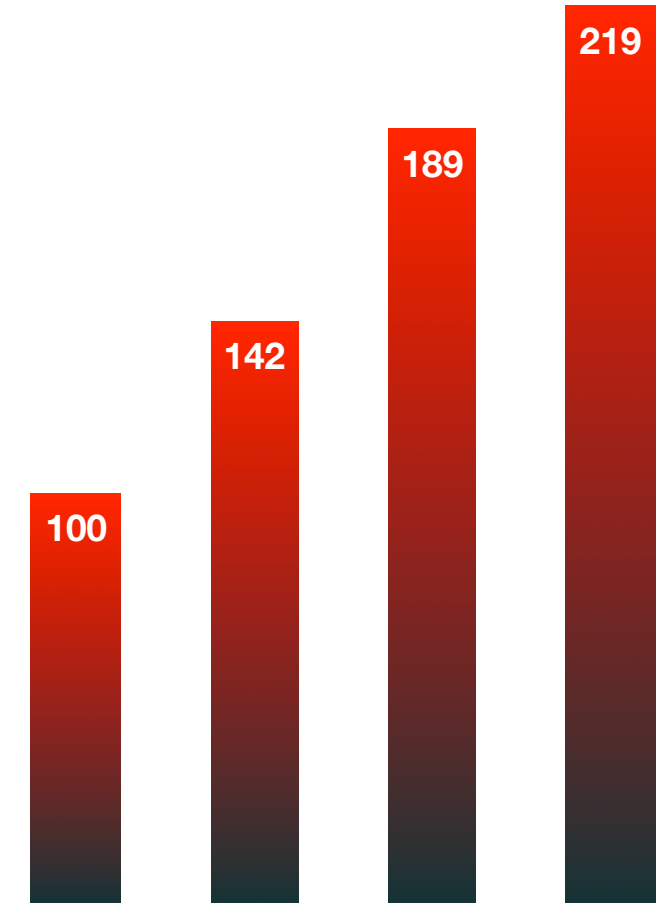
In just three years:

★ total use more than doubled

★ ... **at a staggering compound annual growth rate (CAGR) of 21.7 per cent per annum**.

The growth in article downloads

n=67 UK universities



Source: Sconul / COUNTER 2008

Investment in e-journals, use and research outcomes

Phase II interim findings

Researcher information behaviour

Researcher referencing behaviour

E-journals in teaching and learning

Return on investment

Phase II: Research information behaviour

Investment in e-journals, use and research outcomes

Interview findings (*n*=60)

- ★ very high levels of satisfaction with access to the journals literature.
- ★ at least 95 per cent of journal consumption is now in e-form, with hard copy used only as a last resort.
- ★ a widespread recognition that `nobody' reads `full text' any more, `power browsing' is mainstream.
- ★ Strong preference for generic gateway services such as PubMed and Web of Science rather than publisher platforms.
- ★ Google used to help generate ideas (serendipity) and to fill in gaps: it was felt that systematic reviews and journal articles often omit useful information, the genre is `too rigid'.
- ★ Wikipedia widely accepted as a valuable tool, especially for clarifying terminology across disciplines.
- ★ abstracts are still used to form relevance judgements, but many people prefer to scan the article rapidly, assessing relevance mainly in terms of methodology or terminology (need for structured abstracts?).
- ★ users have little idea how output from services such as Web of Science are ordered and are quite happy to browse through lists of 500 documents.
- ★ Journal brand is still a key quality marker and users actively avoid `inferior titles' and those articles that are poolrly written.

Investment in e-journals, use and research outcomes

Interview findings (*n*=60)

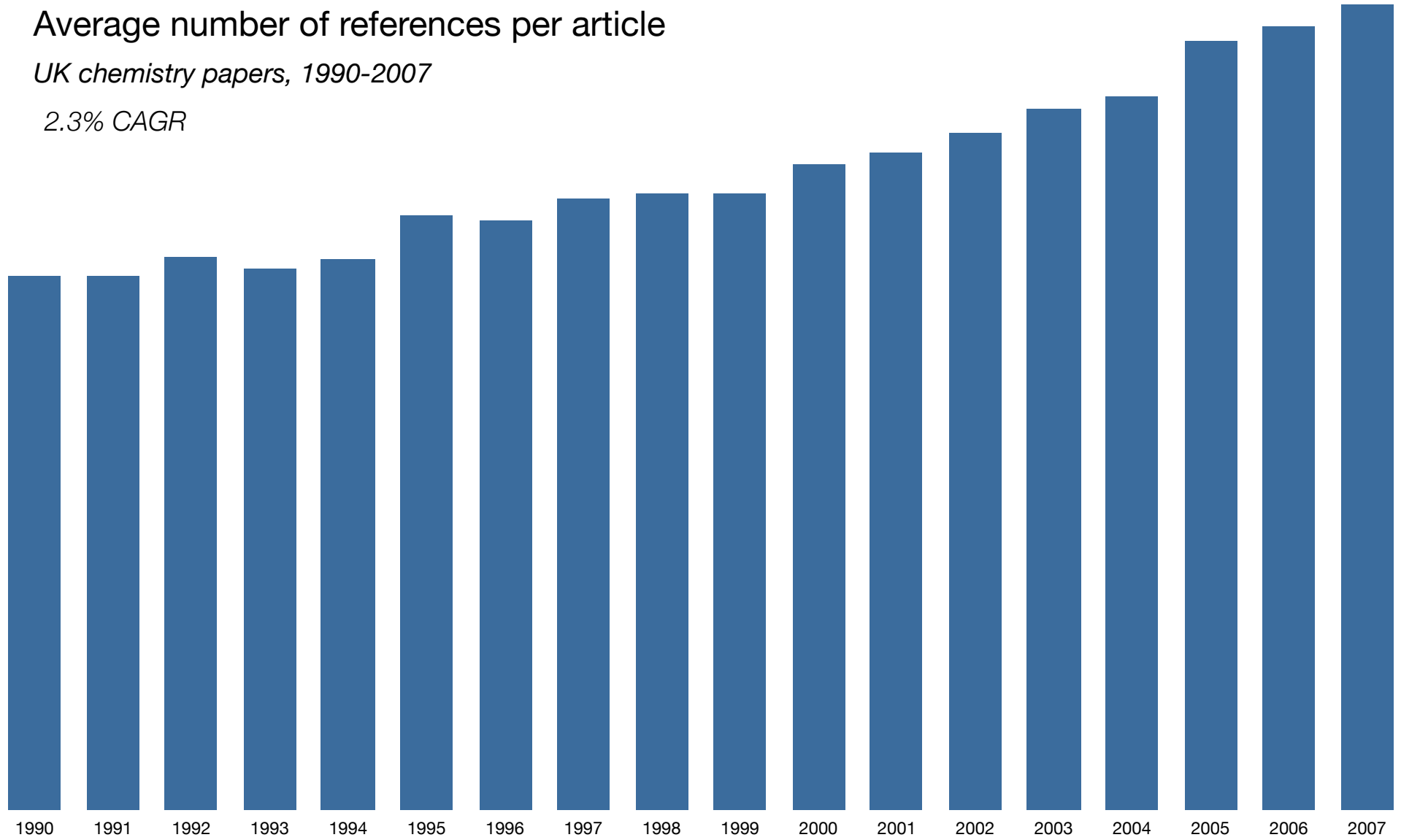
Scientists have always strived to avoid unnecessary reading. Like all researchers, they use indexing and citations as indicators of relevance, abstracts and literature reviews as surrogates for full papers, and social networks of colleagues and postgraduate students as personal alerting services. The aim is to move rapidly through the literature to assess and exploit content with as little actual reading as possible. As indexing, recommending, and navigation has become more sophisticated in the online environment, these strategic reading practices have intensified.

Allen H. Renear and Carole L. Palmer, Strategic reading, ontologies, and the future of scientific publishing, *Science,* 14 August 2009, pp 828-832.

Phase II: Researcher referencing behaviour

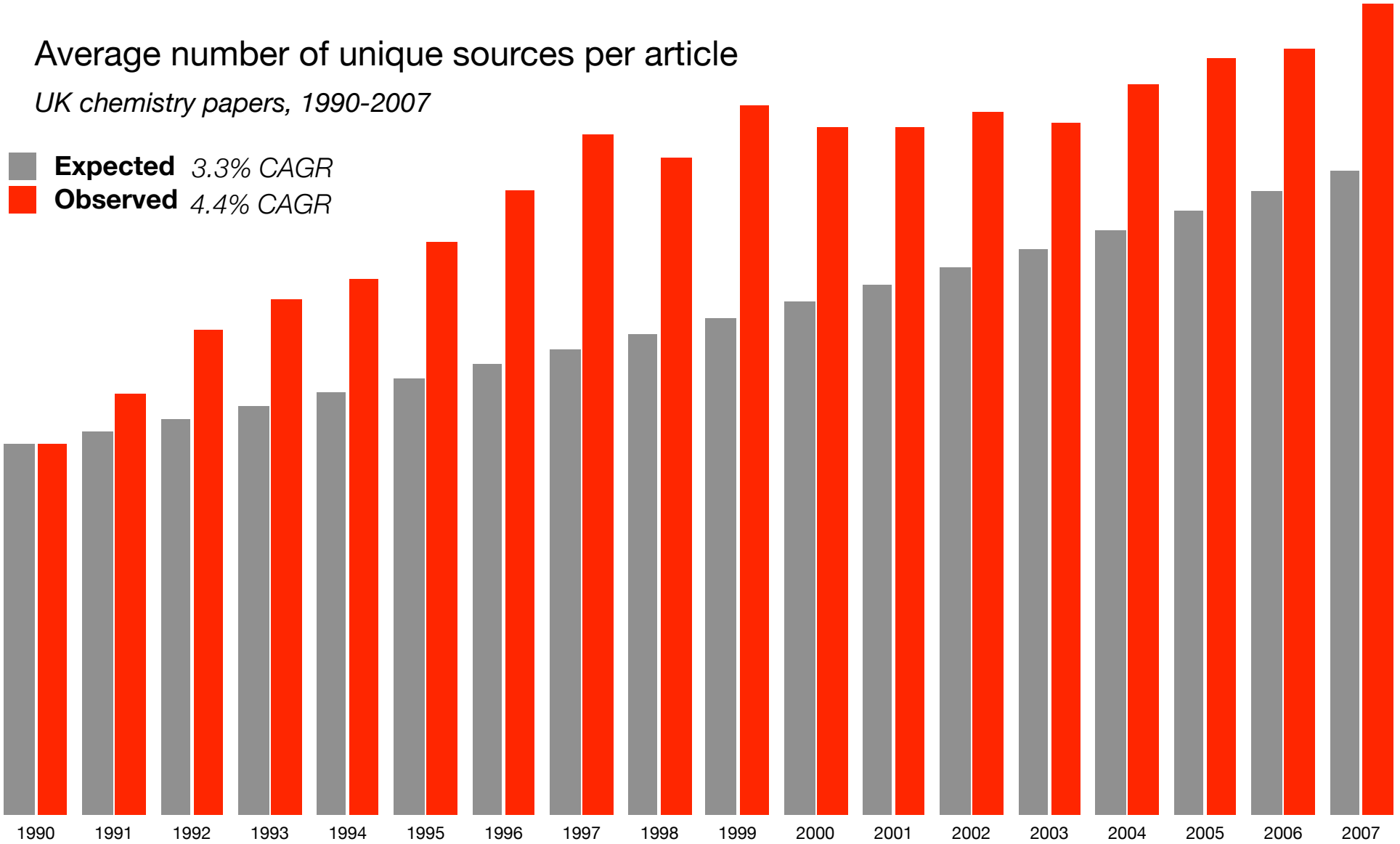
Investment in e-journals, use and research outcomes

Referencing behaviour



Investment in e-journals, use and research outcomes

Referencing behaviour



Investment in e-journals, use and research outcomes

Referencing behaviour

- ★ Authors are citing from more sources now than they did in 1990 and this effect is particularly marked in chemistry and in the earth and environmental sciences. This growth is occurring at a faster rate than that of the literature as a whole. This growth is stronger in the UK than for the world.
- ★ The age profile of references is becoming older.

As a result, researchers are covering the literature in greater breadth and depth, possibly a result of greater access to the literature and new better discovery tools?

Phase II: E-journals in teaching and learning

Investment in e-journals, use and research outcomes

E-journals in teaching and learning

In the RIN Phase I report we wrote:

"It has not been possible to distinguish between use by students and faculty from the publishers' logs on this occasion, but on the basis of published survey findings [by Carol Tenopir and colleagues] we believe that use by undergraduate and Masters' students accounts for around 20 percent of the total. We will be returning to this important issue in the second phase of the study."

Investment in e-journals, use and research outcomes

E-journals in teaching and learning

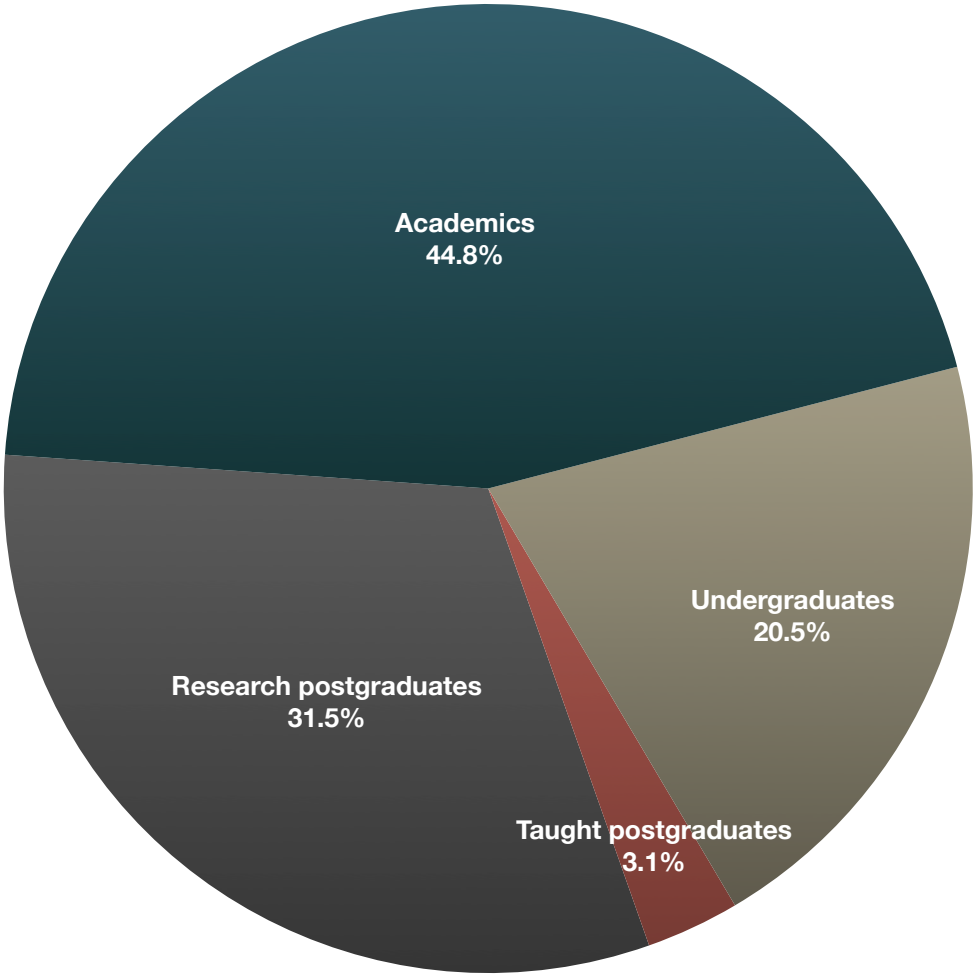
Interview findings

- ★ Journal use in support of teaching is extensive and is mainly about lecturers keeping abreast of their subject in order to prepare classes better and field student questions with greater authority.
- ★ Many lecturers introduce students to the journal literature in their third year, but this is usually very directed. The interviews found that many of these students are surprisingly well informed and knowledgeable.
- ★ There is a general consensus that Masters' students are capable of using that literature independently.
- ★ There is a paucity of e-book compared with e-journal content and it is easily discoverable by students, and e-resources offer concurrent use, unlike print.

Investment in e-journals, use and research outcomes

E-journals in teaching and learning

Estimated share of downloads by academic status 2003/04 to 2007/08
Weighted least square regression model (*n*=116 universities, R^2 = 0.679)



	Full text downloads per annum
Research postgraduates	319.1
Taught postgraduates	10.2
Undergraduates	13.2
Academic staff	314.6

Students possibly account for 23.6 per cent of all e-journal use in UK universities.

`Researchers' (academic staff and postgrad research students) for 76.3 per cent.

Phase II: Return on investment

Investment in e-journals, use and research outcomes

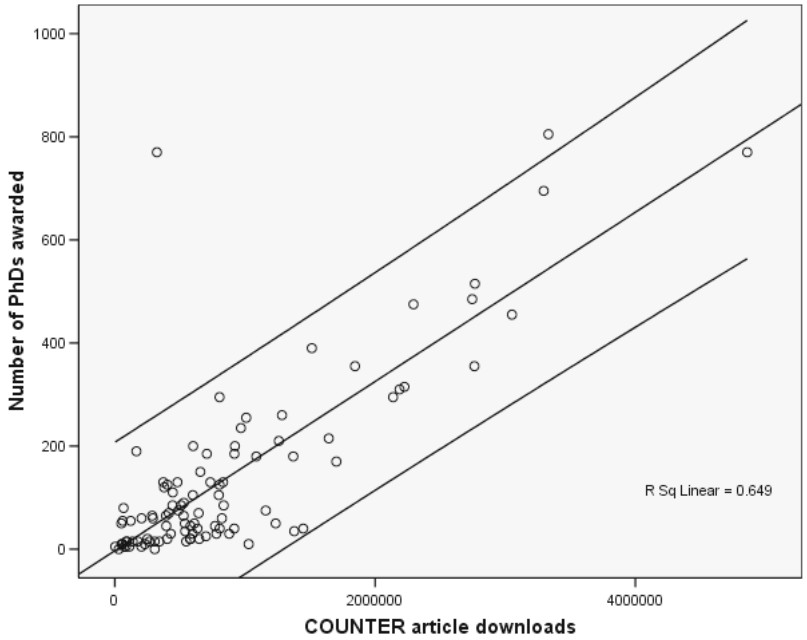
Links between investment, use and `downstream' outcomes?

This slide demonstrates a strong statistical associations between journal use and research outcomes.

The diagram plots numbers of article downloads (as recorded by institutions using the COUNTER standards) against PhD awards for 2006/07. The outer lines are 95 per cent confidence intervals.

The model shows a good fit with few outliers.

PhD awards and article downloads, 2006/07
n=112 UK universities

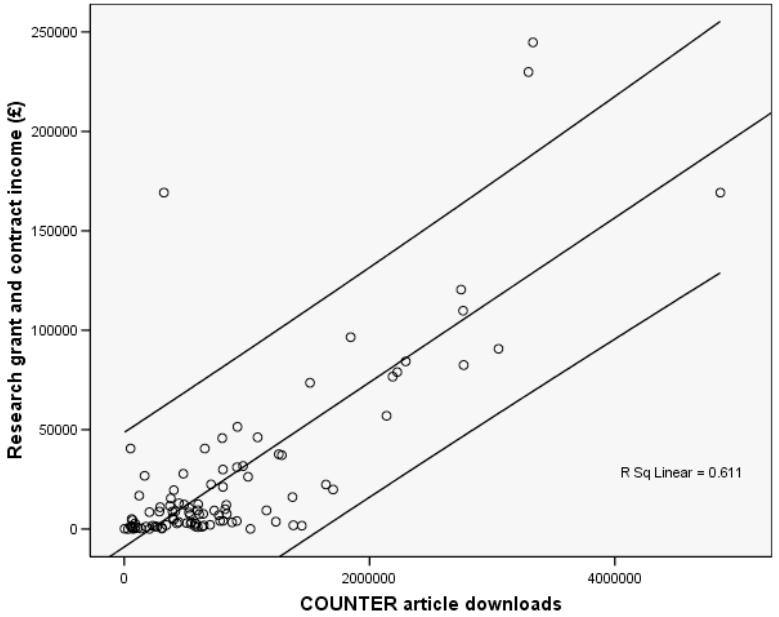


Investment in e-journals, use and research outcomes

Links between investment, use and `downstream' outcomes?

The model shows a good fit with few outliers.

Research grant income and downloads, 2006/07
n=112 UK universities

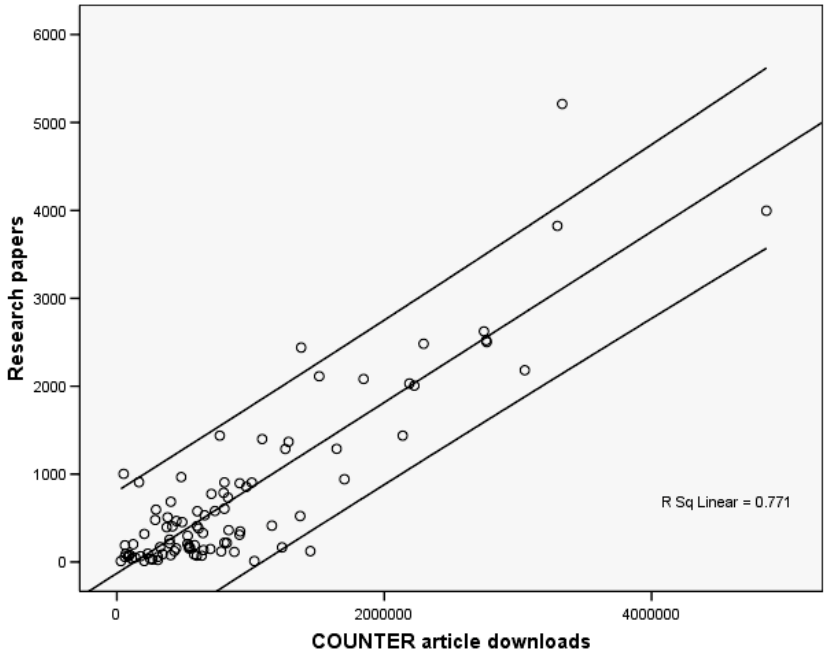


Investment in e-journals, use and research outcomes

Links between investment, use and `downstream' outcomes?

The model shows a good fit with few outliers.

Papers published and article downloads, 2006/07
n=112 UK universities



Investment in e-journals, use and research outcomes

Links between investment, use and `downstream' outcomes?

It is possible to identify three groups of universities on the basis of their downloading behaviour: moderate, high and super users. In the table below, we match these useage groups with measures of their success in research:

	Moderate users (*n*=80)	**High users** (*n*=25)	**Super users** (*n*=10)
Research papers per academic	0.4	0.8	1.0
Research grants and contracts per academic (£000s)	12.7	29.0	39.7
PhD awards per 100 academics	9.1	17.5	17.4
Cost per download	£0.89	£0.74	£0.60

The evidence provided here suggests **a tentative link between e-journal consumption and research outcomes.** Although the differences between the three groups are statistically very significant with respect to research outcomes, there is no such difference in terms of cost per download.

Investment in e-journals, use and research outcomes

Further data modelling

Environmental indicators

- ★ university size measures
- ★ university sector
- ★ research-active staff, subject profile, etc.

Investment indicators

- ★ serials expenditure
- ★ total spend on information content, etc.,

Consumption indicators

- ★ Sconul COUNTER statistics
- ★ ScienceDirect downloads
- ★ loans, e-book accesses, etc.

Output indicators

- ★ article production
- ★ PhD awards

Outcome indicators

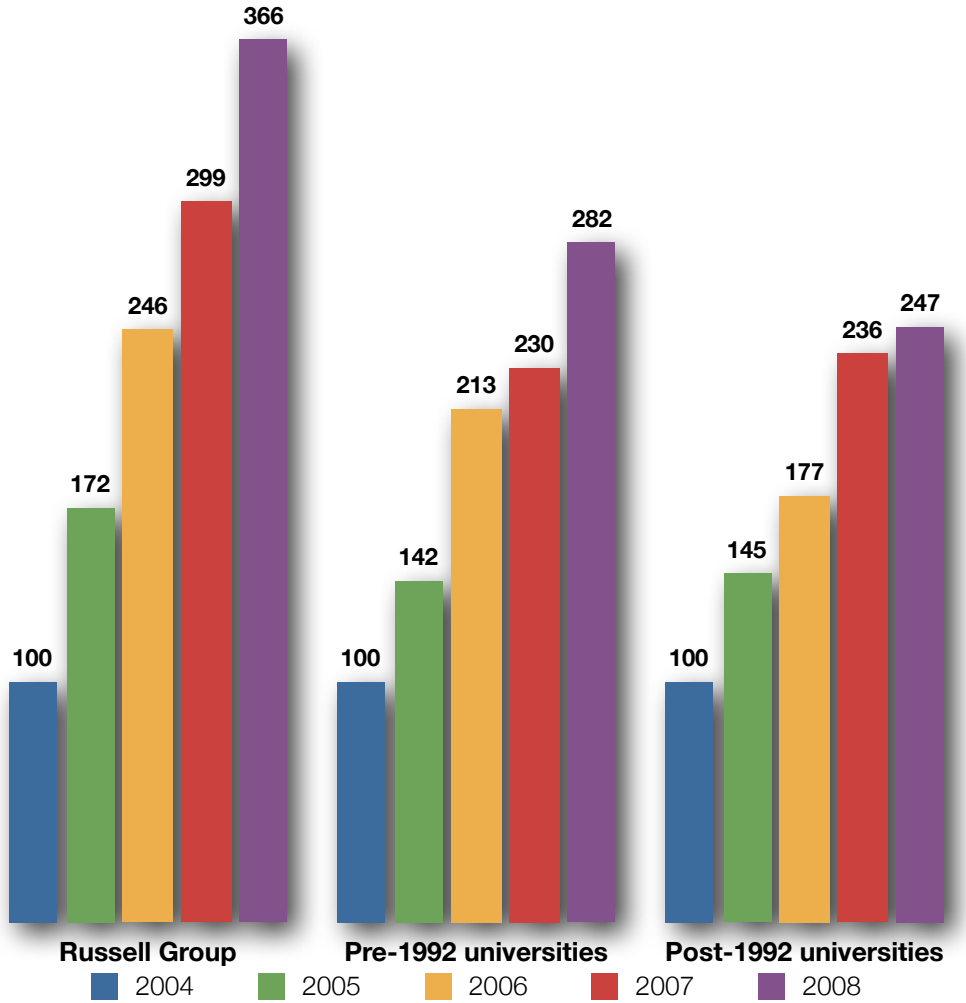
- ★ institutional citation impact (against world average)
- ★ research grants and contracts income
- ★ RAE scores, etc.

Investment in e-journals, use and research outcomes

Further data modelling

There is huge growth in all sectors, but what explains the exceptional growth in download activity by members of the Russell Group? The RAE??

Growth in COUNTER downloads by university sector
(Units per registered library user, indexed 2004=100)



Investment in e-journals, use and research outcomes

Blah

In the first phase of the RIN e-journals research, we established that e-journal investment, use, and research outcomes for 2007 are highly correlated with one another, even when controlling for university size. The challenge this posed for the next stage of the research was whether it might be possible to demonstrate cause and effect and directionality between these factors. Classic chicken and egg territory.

CIBER located a paper that explores a very similar problem, also using official statistical sources as the base (Meso et al., 2009). We have adopted and extended the method used in this paper. The research uses a technique, partial least squares (PLS) regression and path modelling, which is a predictive technique and "particularly useful when predictor variables are highly correlated" (SPSS manual).

The basic idea is to build a model that predicts each of the constructs above (investment, use, research outcomes) on the basis of the other two and to quantify how good they are as predictors of each another.

We have gone a stage further than Meso et al. (2009) in setting up a series of six testable hypotheses, so that we get more of a sense of directionality, if any.

This experiment uses three constructs (investment, use and research outcomes) represented by seven indicators: total spending on p- and p-e journals; total COUNTER downloads, total PhD awards, total RCG income, numbers of articles, and field-normalised world citation impact. The Open University is excluded from this report, leaving 113 institutions.

Indicators from 2005 and 2007 were used so that the models could include a two-year time lag. In the experiment, 2005 independent variables are used to predict 2007 dependents.

The analysis was initially run in PASW v.18 (formerly SPSS) but we concluded that the features offered by a more specialist package, SmartPLS, vastly outweighed PASW in respect of being able to visualise the path models more easily.

Reference

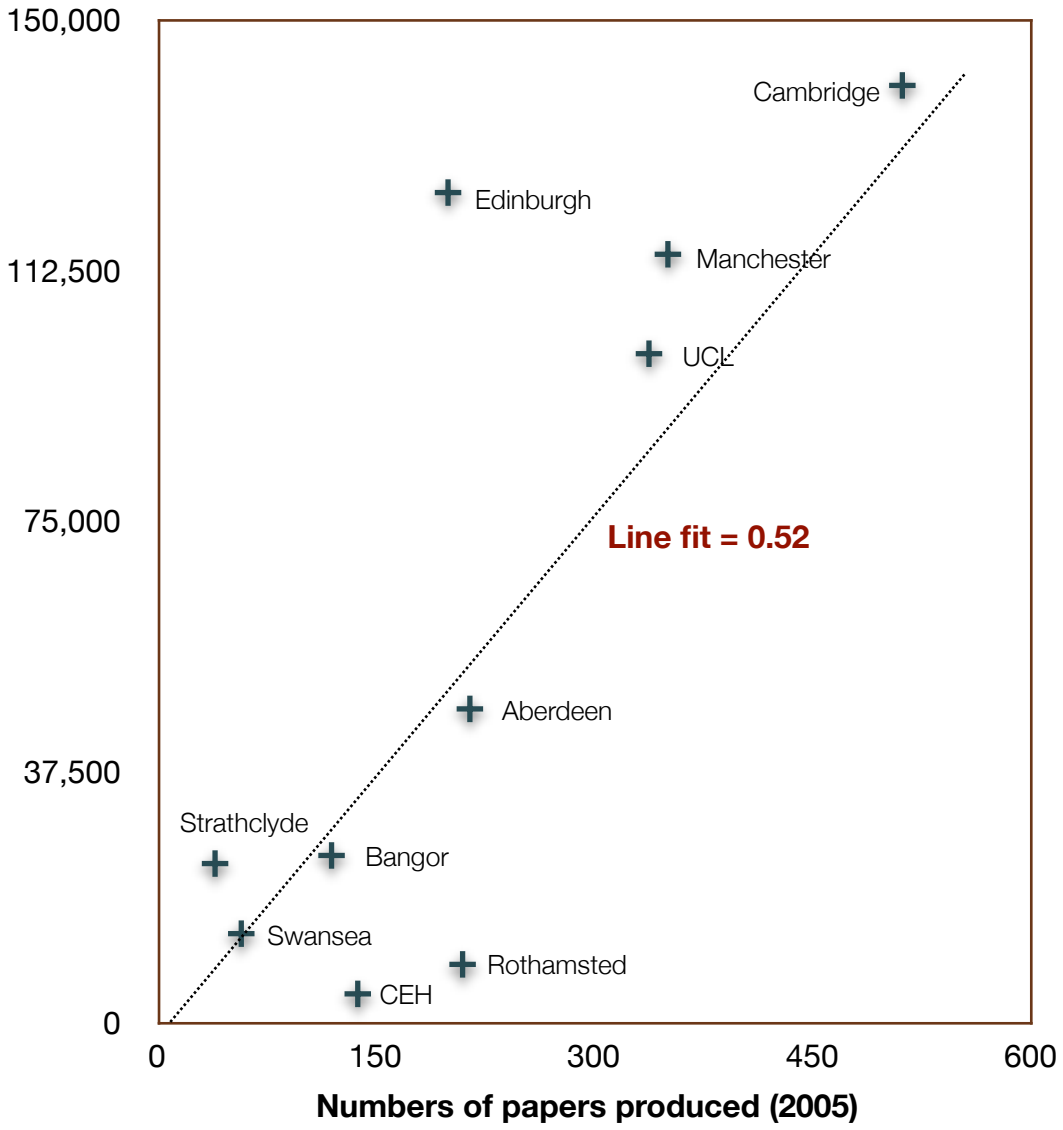
Investment in e-journals, use and research outcomes

Deep log analysis

Life sciences

The evidence here suggests that page views and numbers of articles published correlate reasonably well at the institutional level.

Does this mean that information consumption and production are closely inter-twined?



Investment in e-journals, use and research outcomes

Deep log analysis

The table opposite, for all case study institutions and subjects, shows a number of as yet unexplained but interesting statistical correlations between various information behavioural traits and the Hirsch index (a composite measure of research productivity and citation impact). These correlations are statistically significant but there is much scatter and they only show moderate fit to our computer models. Nevertheless, they are very intriguing and demand further investigation.

The most successful researchers tend to use gateways rather than the publishers' own platforms as their preferred means of information discovery. **Their sessions are shorter and more focused, with fewer pages articles and titles viewed.**

The $64,000 question is does this reflect good practice in the use of these resources, or are there other factors at play?

Information behaviour	Hirsch index
session length	-0.66***
gateway access	0.27*
basic search	-0.52***
advanced search	-0.49**
titles viewed	-0.68**
pages viewed	-0.56***
titles viewed	-0.51**

*** significant at the 1% level
** signiificant at the 5% level
*significant at the 10% level

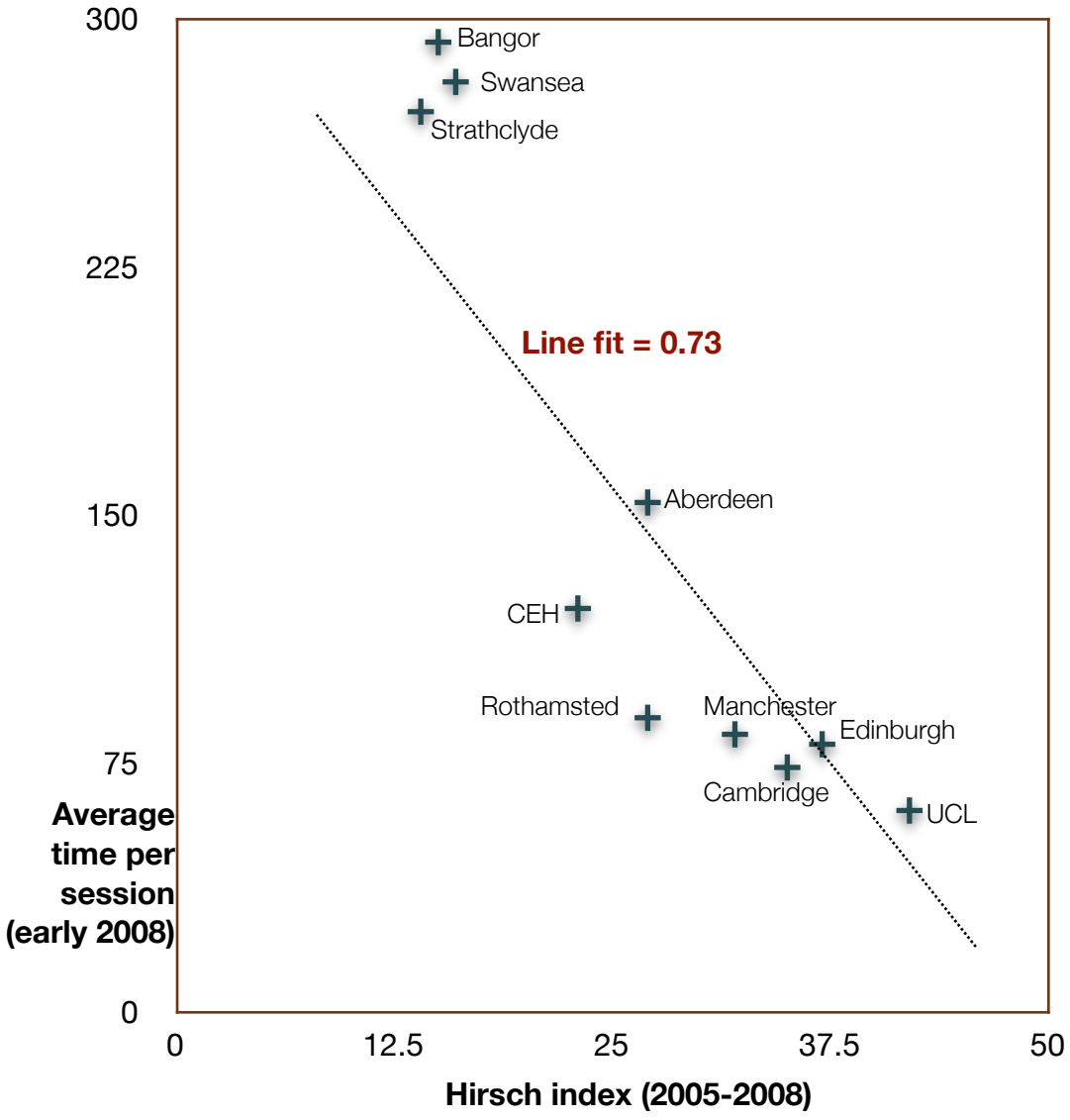
Investment in e-journals, use and research outcomes

Deep log analysis

Life sciences

The evidence here suggests that researchers in the life sciences in more highly-rated institutions average shorter session times (measured in seconds).

Is this because they are `better searchers'? We need to find out.



Investment in e-journals, use and research outcomes

Deep log analysis

A number of strands are beginning to come together as we look at the detailed information-seeking behaviour of our case study institutions and reflect on what these mean for the questions that frame this study.

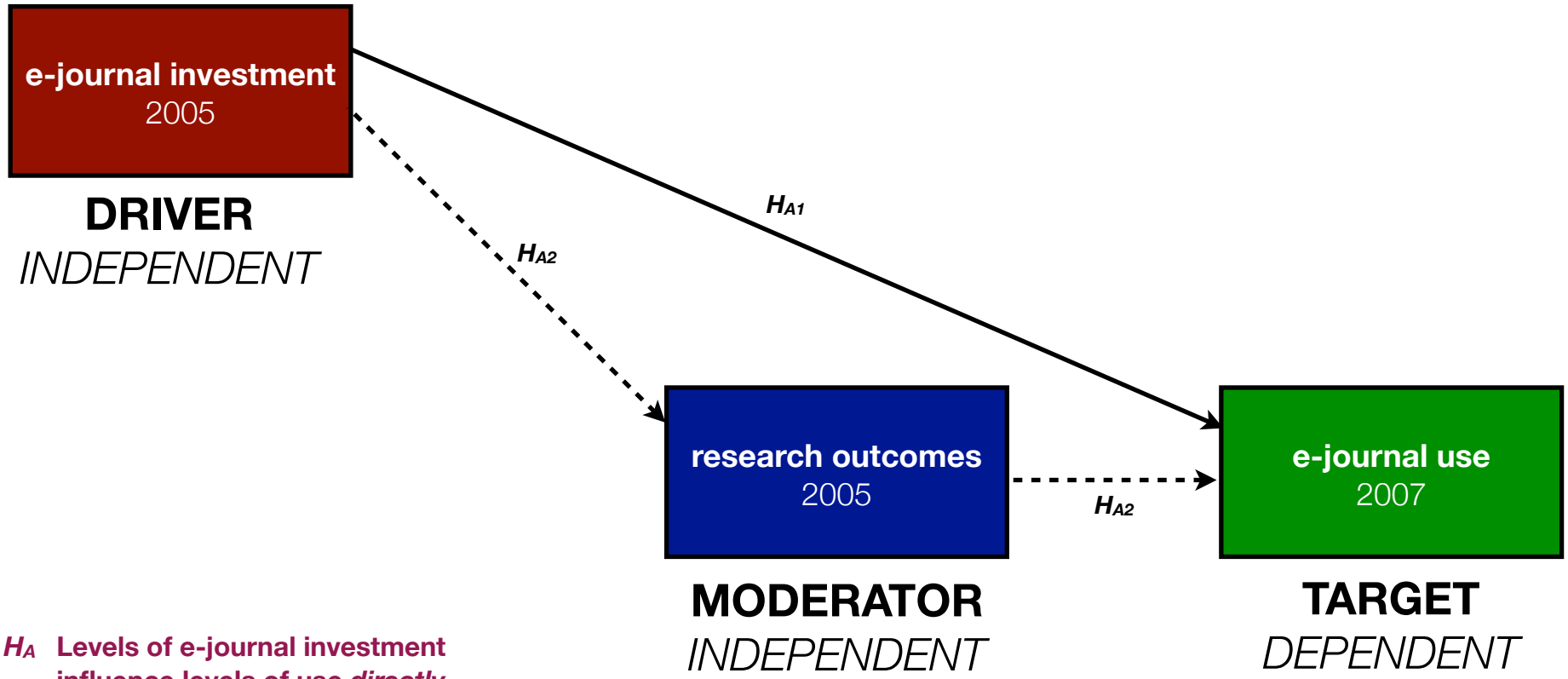
So, for example, we find evidence that:

- ★ information consumption (as measured by page views) correlates with successful publication outcomes;
- ★ researchers in the most productive and successful institutions average less time per session than their colleagues in other institutions when using electronic journals and seem to be more focused in their viewing behaviour;
- ★ researchers in the most productive and successful research institutions consume higher quality information, as measured by the average citation impact factor of the materials they view.

Further evidence of links between journal use and research outcomes is developed in the remainder of this study.

Investment in e-journals, use and research outcomes

Blah



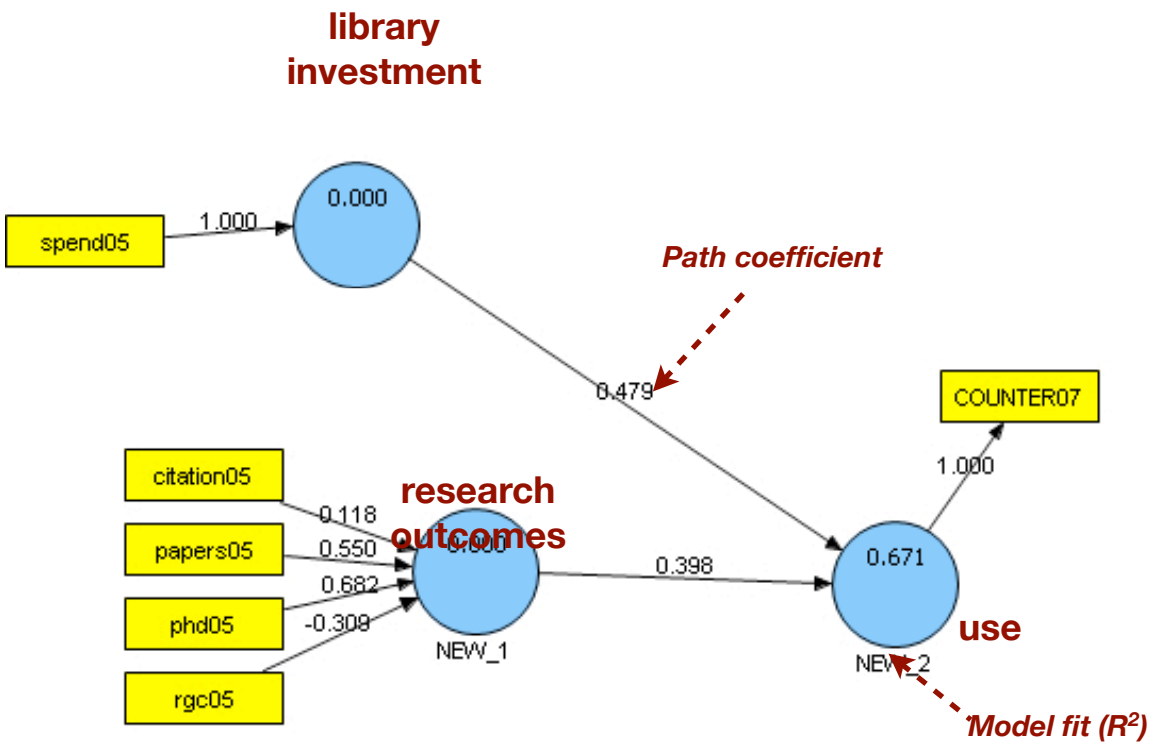
H_{A1} **Levels of e-journal investment influence levels of use *directly***

H_{A2} **Levels of e-journal investment influence levels of use *indirectly* via success in research outcomes**

RESEARCH QUESTION 1
What drives e-journal use?

Investment in e-journals, use and research outcomes

Blah



This path diagram was generated by SmartPLS.

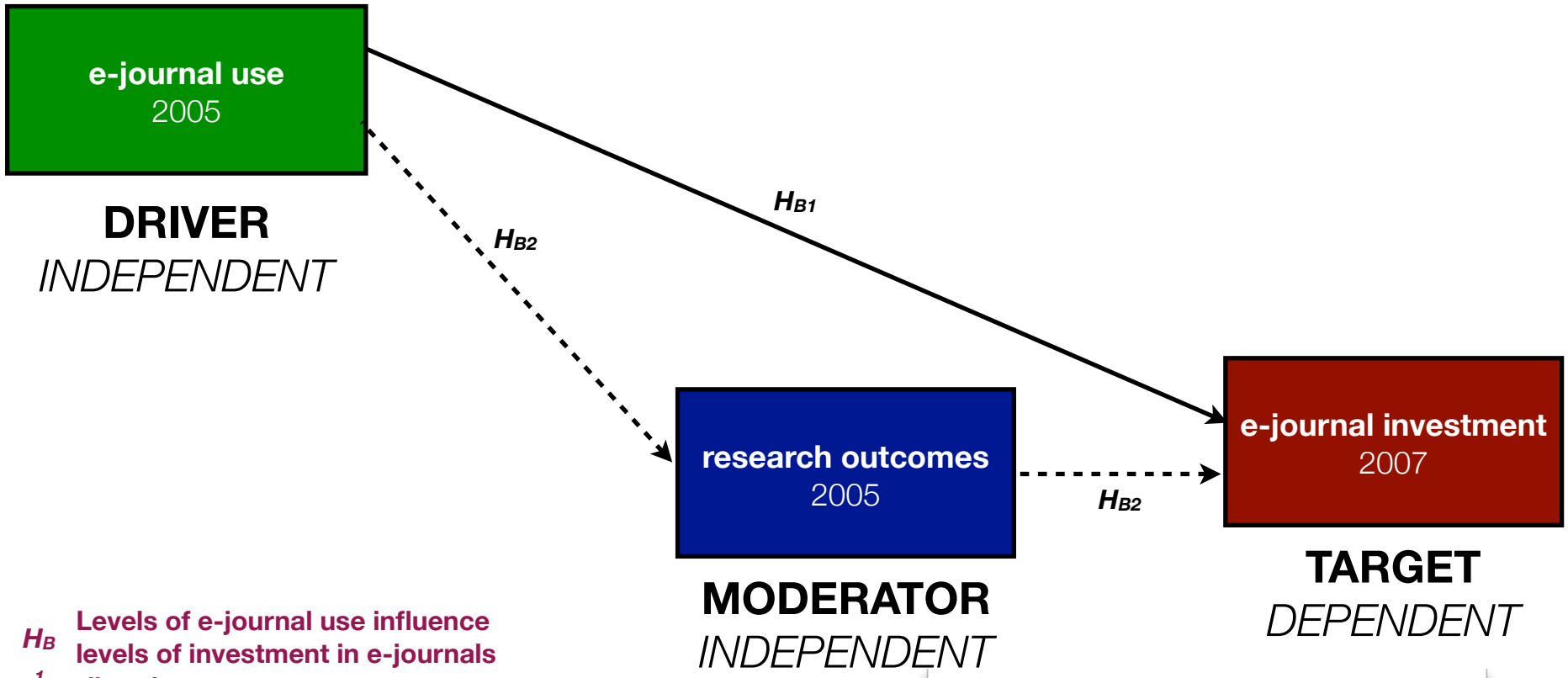
It is a good model: previous levels of library investment and research success are strong predictors of subsequent use (Model R^2=67.1%, the figure inside the bottom `use' bubble).

The path coefficients (ß in the regression equation) show the contribution that the two constructs (earlier investment and research success) make in terms of their predictive ability.

Use is driven more or less equally by previous investment and research success. Key drivers are numbers of papers produced and numbers of PhD students. In this model, RGC income actually has a negative impact on journal use.

Investment in e-journals, use and research outcomes

Blah



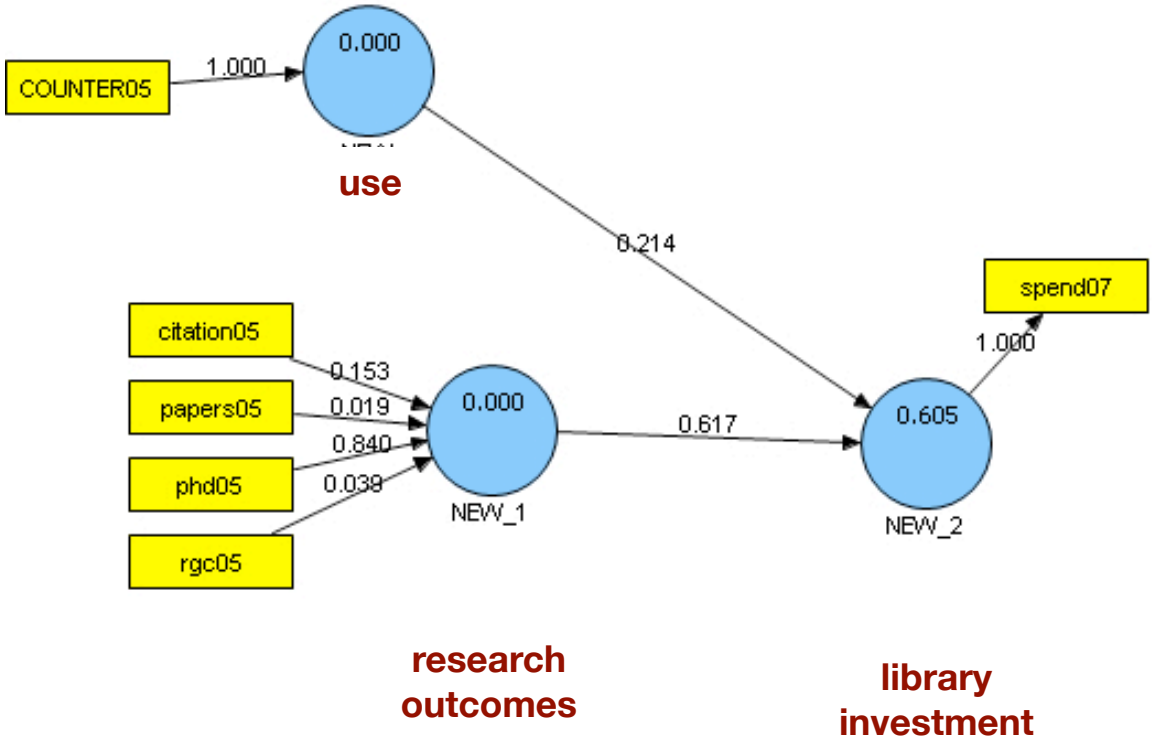
H_{B1} Levels of e-journal use influence levels of investment in e-journals *directly*

H_{B2} Levels of e-journal use influence levels of investment in e-journals *indirectly* via success in research outcomes

RESEARCH QUESTION 2
What drives e-journal investment?

Investment in e-journals, use and research outcomes

Blah



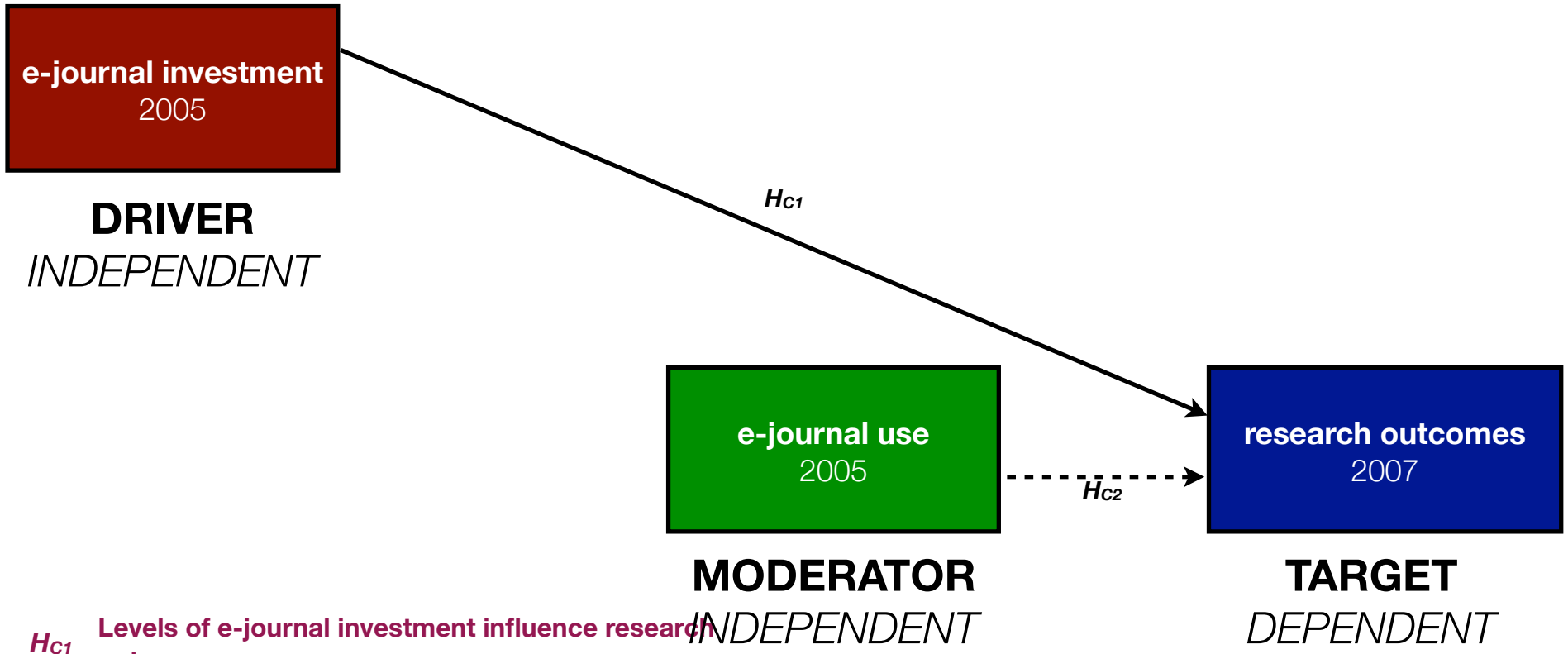
Another good model that shows that previous levels of use and research success are very effective predictors of subsequent investment (R^2=60.5%).

Library investment appears to be driven more by previous research success (a money generator for th einstitution) than by e-journal demand.

Note that numbers of PhD students weigh very heavily in this model.

Investment in e-journals, use and research outcomes

Blah



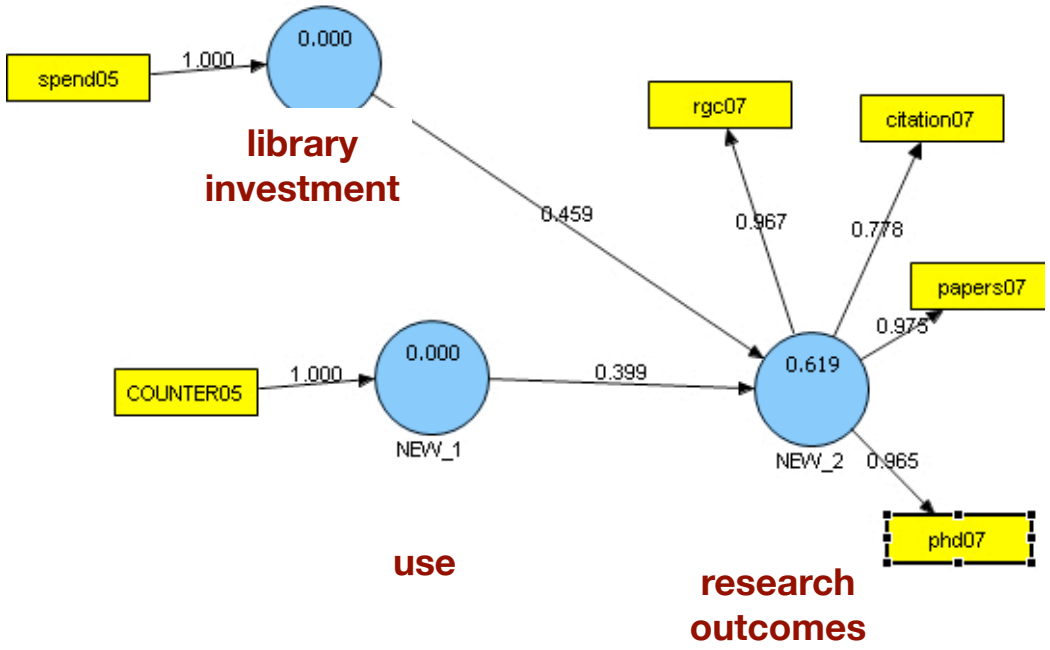
H_{C1} **Levels of e-journal investment influence research outcomes**

H_{C2} **Levels of e-journal use influence research outcomes**

RESEARCH QUESTION 3
What drives research outcomes?

Investment in e-journals, use and research outcomes

Blah



This model shows that library investment and use provide a good model of subsequent research success (model fit R^2=61.9%) and this model is an excellent predictor of specific research outcomes.

On the next slide, we formally test the six hypotheses.

Investment in e-journals, use and research outcomes

Blah

Hypothesis	Path coefficient	T-statistic	*p*	Is hypothesis supported?
H_{A1}: Levels of e-journal investment influence levels of use	0.479	4.147	$p < 1\%$	✓
H_{A2}: Research outcomes influence levels of use	0.398	3.465	$p < 1\%$	✓
H_{B1}: Levels of e-journal use influence levels of investment in e-journals	0.214	1.7932	$p > 5\%$	✗
H_{B2}: Research outcomes influence levels of investment in e-journals	0.617	4.460	$p < 1\%$	✓
H_{C1}: Levels of e-journal investment influence research outcomes	0.459	2.030	$p = < 5\%$	✓
H_{C2}: Levels of e-journal use influence research outcomes	0.399	1.779	$p > 5\%$	✗

The criteria used for accepting a hypothesis is a path coefficient greater than or equal to 0.2 and a T-statistic greater than or equal to 2.02 (the threshold t-statistic for significance at the 5% level for five degrees of freedom see Meso et al., 2009, p.59).

Two of the hypotheses above are not supported at the 5% level.

Investment in e-journals, use and research outcomes

Blah

This analysis of 113 universities draws the following provisional conclusions:

Previous investments in e-journals are a strong predictor of future use (HA_1), and this finding suggests that library journal budgets should at least be held at current levels and not cut in real terms.

Previous success in research is also a strong predictor of future e-journal use (HA_2). Together with the previous finding, this suggests that usage is strongly reinforced by research culture and activity and high levels of library spend. This will be of surprise to no one, but here is firm evidence.

The evidence that use is a direct driver of library investment in e-journals (HB_1) in this experiment is weak (this is unsurprising since journals are generallly supplied as bundled deals and not priced according to usage).

The most powerful relationship shown here is that previous research success drives library investment going forward (HB_2). This is highly plausible since research success generates additional postgraduate fee income and research overheads for the institution.

Library investment in e-journals is a strong driver of subsequent research outcomes (HC_1).

There is no strong evidence of a direct `causal' (i.e. predictive) link between aggregate levels of COUNTER downloads and research outcomes. A complicating factor here is that we do not know the proportion of use that should be attributed to research and teaching.

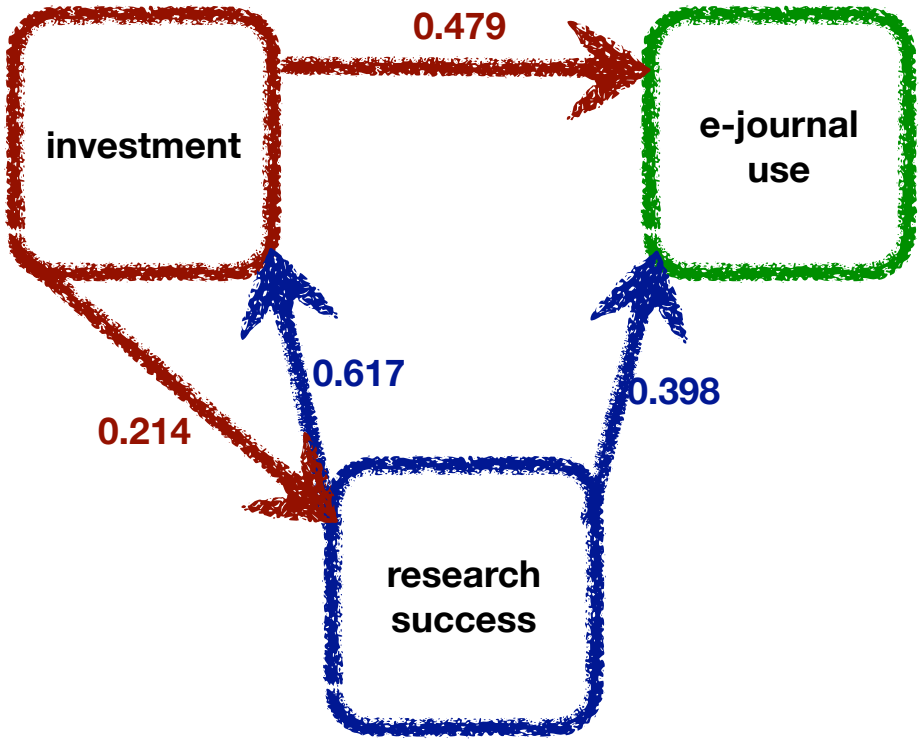
Investment in e-journals, use and research outcomes

Blah

This diagram summarises the previous discussion. It would appear that levels of research activity and library investment drive use, not the other way around, It would also appear that there is some evidence of directionality between library investment and research success. A possible interpretation is that research success generates additional income for the university and that this is the major driver here.

İLimitations

This is a small-scale study to explore the concepts behind path modelling to see if they could be used in studies of library return on investment. The main challenge for taking this forward is that we have no obvious means of separating out the use of e-journals for research or for learning and teaching at each institution. As a result, the links between research success and total COUNTER downloads (in support of research, learning and teaching) are tenuous.



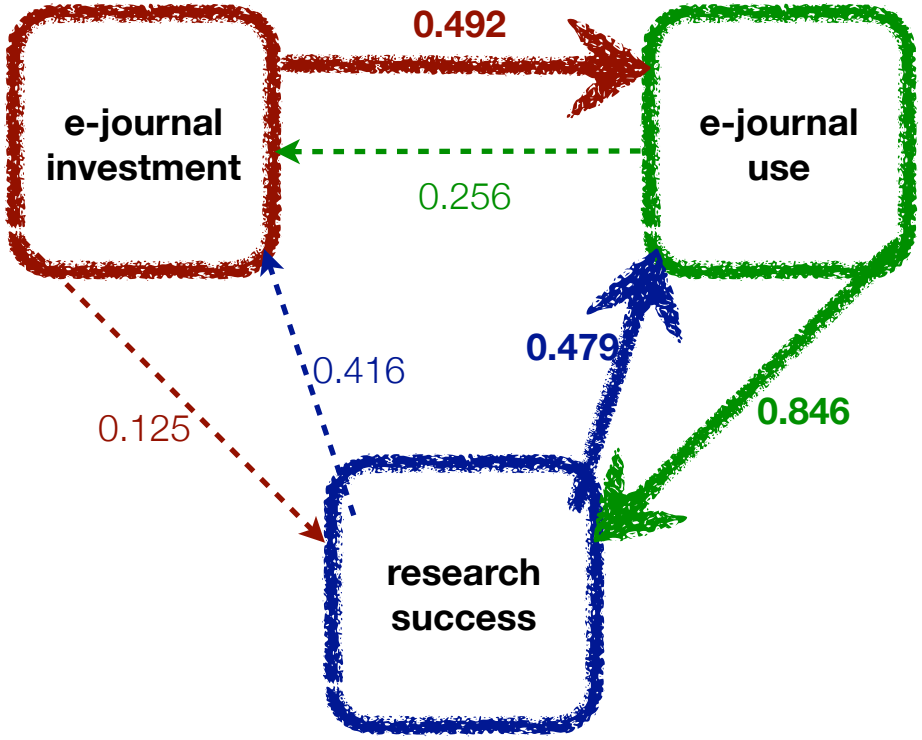
Investment in e-journals, use and research outcomes

Further data modelling

The balance of the evidence is good news for the library community:

use is a strong driver of research success (and creates a positive feedback loop)

continuing library investment is crucial



Annex I

Constructs and indicators 2005 data (*n*=113)

CONSTRUCT	INDICATORS	LABEL	*N*	MIN.	MAX.	MEAN	SD
INVESTMENT	Total expenditure on e- and p-e journals (Sconul)	spend	113	£5,289	£1,924,230	£501,950	438,073
USE	COUNTER downloads (CIBER estimates)	COUNTER	113	180	2,937,855	634,417	574,012
RESEARCH OUTCOMES	PhD awards (HESA)	phd	113	0	920	134.1	177.3
	Articles published (Elsevier Scopus)	papers	113	1	5,905	879.7	1,.236.5
	RGC income	rgc	113	£68	£201,576k	£25,752k	42,082
	Field-independent citation impact (Elsevier)	citation	113	0.15	1.94	1.03	0.37

Annex II

Constructs and indicators 2007 data (*n*=113)

CONSTRUCT	INDICATORS	LABEL	*N*	MIN.	MAX.	MEAN	SD
INVESTMENT	Total expenditure on e- and p-e journals (Sconul)	spend	113	£45,232	£2,889,708	£674,472	628,442
USE	COUNTER downloads (CIBER estimates)	COUNTER	113	388	4,859,439	927,506	847,073
RESEARCH OUTCOMES	PhD awards 2008 (HESA)	phd	113	0	995	148.4	189.9
	Articles published (Elsevier Scopus)	papers	113	2	6,571	990.9	1,363.7
	RGC income	rgc	113	£87k	£251,623k	£28,680k	49,324
	Field-independent citation impact (Elsevier)	citation	113	0.14	1.96	1.09	0.37

Annex III

Pearson correlation matrix 2007 data

	spend	COUNTER	phd	papers	rgc	citation
spend	1.00					
COUNTER	0.68**	1.00				
phd	0.75**	0.67**	1.00			
papers	0.70**	0.71**	0.94**	1.00		
rgc	0.69**	0.70**	0.94**	0.97**	1.00	
citation	0.55**	0.55**	0.64**	0.67**	0.65**	1.00

*** Correlation is significant at the 1% level (2-sided)*

This table (2007 data) shows just how strongly the indicators are correlated one with another. This feature (multicollinearity) limits the number of statistical tests that can be used: most assume a degree of independence between the variables used. Not so PLS modelling.

Key to indicators

spend (Total spend on e- and e-p journals)
COUNTER (Total downloads)
phd (Total PhD awards)
papers (Total articles published)
rgc (Total RGC income)
citation (Field-independent citation impact)