Understanding Barriers to the Use of Numeric Data in Learning and Teaching

by Robin Rice[1]*

Background
UK higher education is rich in numeric datasets. In the socioeconomic field, for example, there are large-scale, representative sample surveys (e.g., General Household Survey), current and historical population censuses, international comparative datasets, longitudinal surveys, economic time series, and data about markets, companies, and commerce. In the UK a centrally funded system of national data services for higher education provides for the dissemination of much of this research data, which is free at the point-of-use and accessible over the Internet (via JANET, the UK academic network).

However, these data resources are under-used in the learning and teaching environment. Despite the potential gain in numeracy, critical use of evidence and empirically-based knowledge by students conducting data analysis at both the postgraduate and undergraduate levels is infrequent, and obstacles exist that make integration of numeric data resources into coursework difficult. Employing numeric data effectively in teaching requires specialised skills and more time for preparation than the use of printed materials or bibliographic databases, and both students and teachers require a high level of support. As expectations about the use of information technology in learning and teaching rise, the barriers that inhibit the use of this wealth of data in the classroom and in student projects need to be lowered.

Understanding statistical evidence is important not just for postgraduates learning to be researchers and entering the professions, but for undergraduates as well. Milo Schield has written widely about teaching statistical literacy in higher education. He explains it as a different and more fundamental skill than producing or 'doing' statistics: "Statistical literacy focuses on making decisions using statistics as evidence just as reading literacy focuses on using words as evidence. Statistical literacy is a competency just like reading, writing, or speaking."[2] The need for application of such a competency in many fields is readily apparent.

The Numeric Data Project
This paper reports findings from a national collaborative project: "Using Numeric Datasets in Learning and Teaching," funded by the JISC (Joint Information Systems Committee, which itself is funded by the Higher Education Funding Councils). The lifetime of the project is February 2000 to September 2001. Project partners are from three national data centres, EDINA, MIMAS, and the Data Archive, and two university data libraries, the University of Edinburgh and the London School of Economics. Additionally, a Task Force of experienced academics from across the UK was recruited as volunteers to guide the enquiry and its outcomes. This partnership reflects the novel perspective taken by the project to examine use of the nationally-funded data services with particular reference to local support needs of teachers and learners within their universities. The project is one of several funded under the JISC's Learning and Teaching Development Programme (see http://www.jisc.ac.uk/dner/programmes/projects/ for a full list of projects).

A major objective of the project was to generate knowledge on issues such as the extent of use and the practicalities of using data in teaching, and the experiences teachers have of data support from both national data services and support staff in local institutions. Since user surveys tend to target those already registered for national services, there is no ready evidence about the larger population of UK university teaching staff on these issues. Therefore, a nationally representative sample survey was needed to discover the current "state of play" before recommendations about how to lower barriers could be made. The survey was designed to ask teaching staff about their use of numeric data in teaching and supervising students, their experience of national data services, barriers to using data in teaching, and the extent of support available within their institutions.

The teachers' survey was enhanced by qualitative case studies of a diverse set of postgraduate and undergraduate classes using numerical data in teaching, which both inform the enquiry and also act as exemplars for other teachers. The full survey results and case studies are available on the project Web site at http://datalib.ed.ac.uk/projects/datateach.html. The final report with its recommendations, teaching resources, and other information is also available.

Survey Methodology
A sample postal survey was conducted of UK university teaching departments within the social sciences, plus other selected disciplines "outside" the social sciences, such as public health sciences. Two hundred sixty-seven department heads were randomly selected from a universe of 1590 (1 in 6 sampling fraction). The sampling frame was purchased from the marketing company Mardev, extracted from the *Worldwide Academic & Library File*. Department heads were asked to complete the four-page questionnaire themselves and to pass copies to relevant teaching colleagues to garner their participation. (A Web version was also made available for on-line input.) There were 206 responses collected from 110 departments. Fifteen records were removed as ineligible (e.g. non-teaching department). Following telephone, e-mail, and postal follow-up requests to sample members, the final response rate (110 / 252) was 44 percent of departments sampled.

Survey Results: Use of Data in Teaching and Learning
Due to the survey design and instructions to department heads, there was likely a skew toward data users among those in the sample who participated, as a result of self-selection. (Non-data users tended not to respond to the survey, as it was not felt to be relevant to them.) Seventy-nine percent of those survey respondents who taught or convened courses used data either "nearly always," "often," or "occasionally" (see Chart 1). The sample also seemed to over-represent senior staff (perhaps because the request was sent to department heads), teachers of methods courses, and those committed to quantitative analysis. Among those who used numeric data in teaching in some form, about two-thirds expected students to work with data
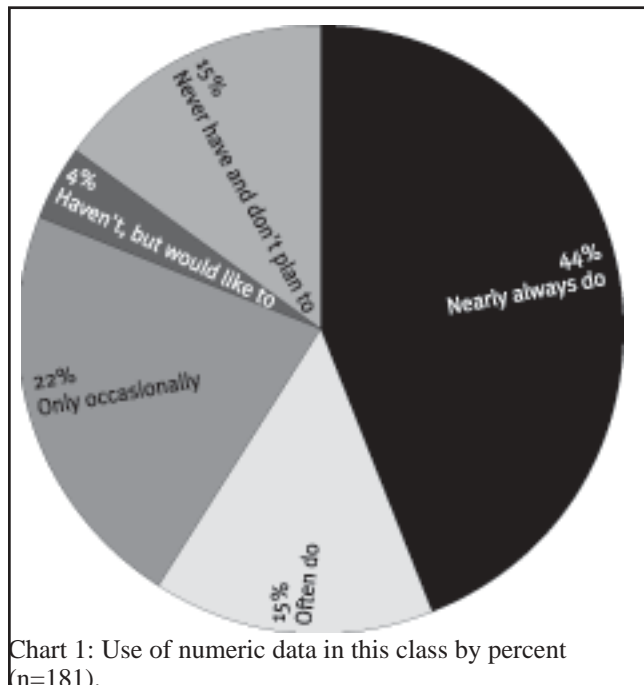

Chart 1: Use of numeric data in this class by percent (n=181).

on a computer, in "hands-on" fashion. As Table 1 shows, a higher proportion of methods courses were hands-on than subject courses. [The categories of "methods-based" and ""subject-based" were coded during analysis, based on names of courses supplied by respondents.] Surprisingly, neither course level nor class size appeared to affect whether the course was hands-on.

col %	Methods	Subject	All
Hands-on	85	54	64
Not hands-on	15	46	36
n=	46	100	146

Table 1: Whether course is "hands-on," by course type

Although the survey was directed towards staff, not students, there was an attempt to understand the level of data use by students in their independent learning. Ninety-two percent of respondents who were either post- or undergraduate supervisors recommended the use of numeric data for students''research at least occasionally (depending on the nature of the research project). Below are "typical" responses for each category.

- *Nearly always do (35 percent):* "Statements made need to be backed up with evidence – often of an empirical nature."

- *Often do (33 percent)*: "Depends on topic, but statistical sources can contextualise a topic."

- *Only occasionally (21 percent):* "Many students are more inclined to qualitative research."

- *Never have and don't plan to (6 percent):* "Not relevant to what I am teaching."

- *Haven't yet but would like to (2 percent):* "Not always appropriate and [I am] insufficiently briefed on numeric data available."

Burden of Data Preparation
The survey instrument dealt directly with the issue of how burdened teachers felt regarding data preparation. As Chart 2 shows, a slight majority felt that data preparation was a burden, but warranted.
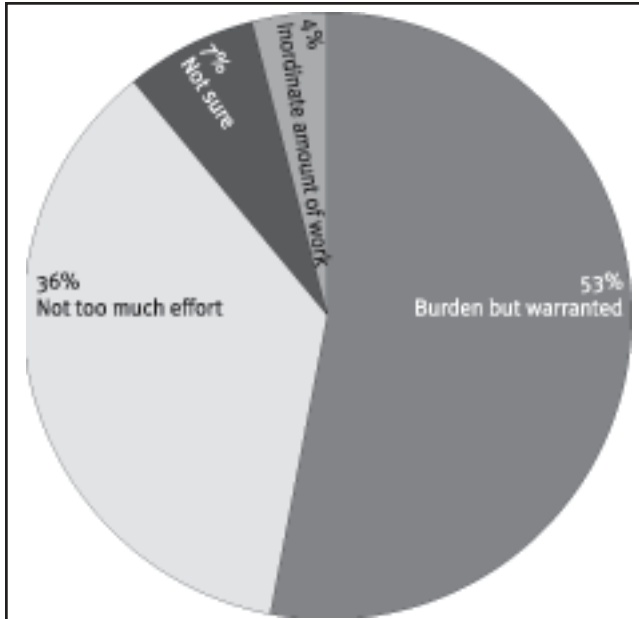
Chart 2: Burden of data preparation, percentage of respondents (n=140

4%
Inordinate amount of work

7%
Not sure

36%
Not too much effort

53%
Burden but warranted

serial. (Print publications obviously do not provide the material needed for a "hands-on" component, which gives students practice at manipulating data on a computer, unless the data are hand-entered.) The rest of the sources, including from a colleague, freely available on the Internet, or bundled with a textbook, were used by less than 20 percent of teachers who use data. Twice as many respondents received data from a government agency or "directly from the data producer" as were registered with a national data service.

These results indicate a need to further explore the nature of data sources needed by particular disciplines for teaching particular types of courses, and whether the national data services and local institutions are providing adequate collections. The findings also seem to undermine the notion that anything needed can be obtained freely on the Internet. Financial and company datasets, for example, are profitable information commodities, which require substantial academic discounts or subsidies to be affordable.

Would the national data services be more widely used if they were providing relevant collections to teaching departments? A closer look at the barriers to use of the national data services uncovers deeper issues than just ensuring that available sources exist.

Respondents were also asked if they felt the need to update / refresh / revise the data used on a regular basis. Of those responding (78 percent of those eligible, n=142), 57 percent said yes, and only 14 percent said no. However, 29 percent said yes, but there was insufficient time to do so.

Data Sources and Use of National Data Services
The survey showed quite clearly that, although the use of numeric data among the survey respondents is high, the use of *national* data services that provide on- or off-line access to secondary datasets is not. ***Only one-quarter*** of the respondents who used data in teaching had "used or considered using" the national academic data services (namely the Data Archive, EDINA, and MIMAS) for teaching purposes.

So what are the sources of numeric data used in higher education classes? Most strikingly, half the teachers either required their students to collect their own data, or taught with data they collected themselves (see Chart 3). Nearly half, 44 percent, used print data sources, extracted from a monograph or
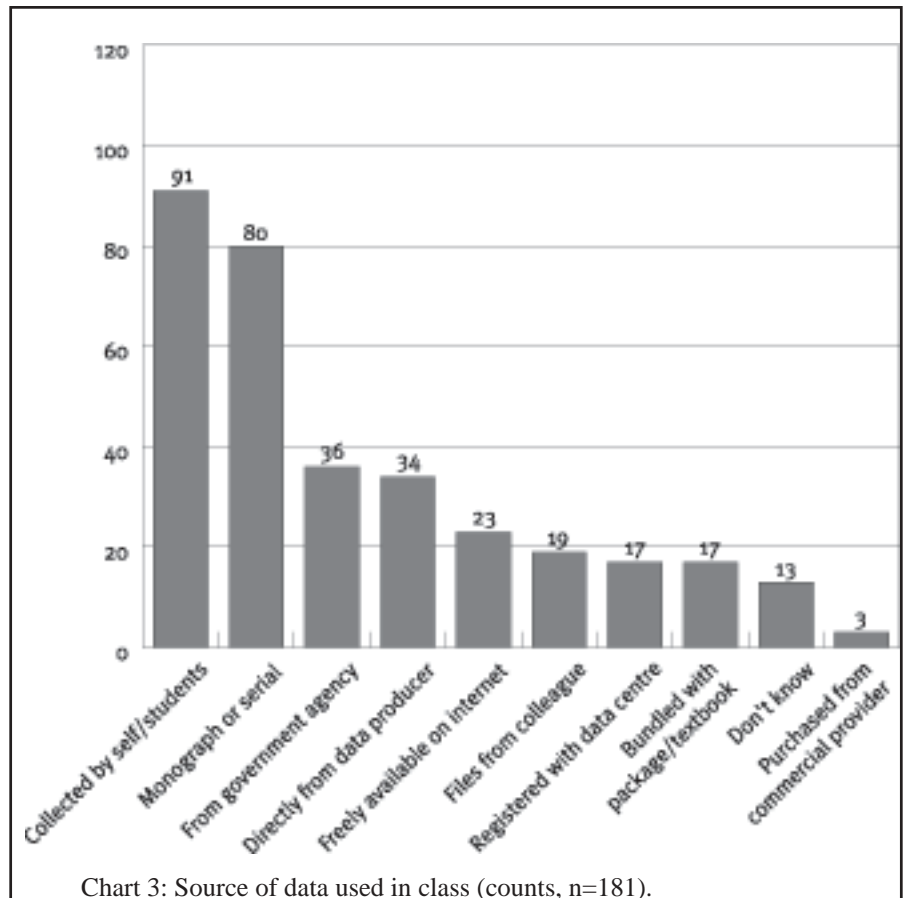

Chart 3: Source of data used in class (counts, n=181).

Barriers to Using Datasets in Teaching

Those 46 respondents who were familiar with the national data services (one-quarter of those who teach with data) were asked to rank eight factors they thought might act as barriers in using national data services for learning and teaching purposes. Table 2 shows the median score for each barrier, in descending order, and also the mean score. The two top-rated barriers were "lack of awareness of relevant materials," and "lack of sufficient time for preparation." This issue was highlighted in a separate question, in which 57 percent agreed on the need to update /refresh /revise datasets used for teaching, but 29 percent had insufficient time to do so. The third greatest barrier was "registration procedures" [of the national data services]. However, the other barriers received high enough scores to also be considered seriously: namely, difficult data extraction interfaces, unsuitable file formats, inadequate dataset documentation, and lack of tailored teaching subsets.

	Mean score	**Median Score**
Lack of awareness of materials	6.5	**7**
Lack of time for preparation	6.4	**7**
Registration procedures	5.6	**6**
Interface	5.0	**5**
Format of datasets	4.8	**5**
Documentation	4.6	**5**
Lack of teaching subsets	4.4	**5**
Table 2: Average ranking of barriers (8=highest score, 1=lowest).		

In an open-ended question, users were asked for positive changes the national services could make to support teachers and learners in the use of datasets. Thirty-six out of 46 eligible respondents answered the question with a variety of useful suggestions. Answers were grouped into the following four categories (with examples of actual responses):

• *Easier access* - Able to get data without learning special software."

• *Simple registration for students* —"Make registration procedures simple and abolish restrictions on use (e.g. all students signing disclaimers)."

• *Create relevant and interesting teaching datasets* – –"Rapid access to key summary economic data in form tailored for teaching."

• *Effective publicity* —"The initiative needs to come

from the National Services but better publicity would be a start."

Support Issues

Prior to the survey, only anecdotal evidence was available to determine how teachers obtained support for classroom use of datasets. Members of the Task Force were familiar with the common reality of peer support for data use in both research and teaching via word-of-mouth. One member was aware that he was considered to be "the data guy" in the department, to whom others came for support. Although two data librarians were involved in the project, specialised data libraries and data librarians are not common in UK universities. Site representatives for the national data services can be based in the library, comput-ing service, or elsewhere in an institution, but it was not known how much support they actually provide to users.

To provide a baseline measure on this issue, the survey asked each respondent,"From whom have you ever had support in obtaining or using data, *whether for teaching or for research?*"Of those who responded, more than a third (37 percent) had received no support at all. More than one source could be ticked; the average number of sources of support received was two. Peer support was the most common form, either from a project co-worker/assistant or another colleague (26 percent and 47 percent, respectively). The local computing service (26 percent) was roughly matched with the local library service (23 percent), which had helped about a quarter of respondents each. National service staff provided help to 10 percent of respondents, and their local site representatives only helped 7 percent of

Chart 4: Level of local support provided, percentage of respondents (n=176).
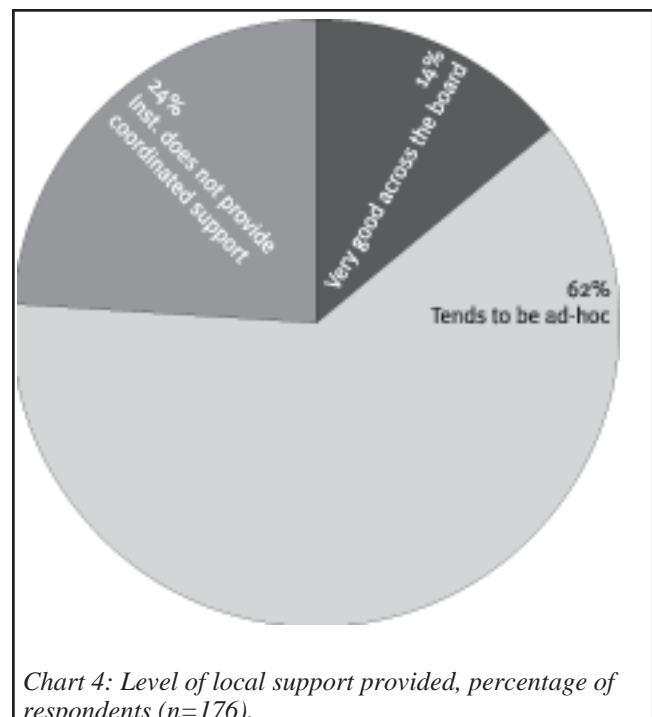
them.

As an indicator of the satisfaction level with this status quo, users were asked to characterise the level of data support provided in their institution. The results are shown in Chart 4. Notably, only 14 percent agreed that local support was "very good across the board." The majority, 62 percent, felt that support "tends to be ad-hoc."

To follow this up, the survey instrument anticipated a number of local support activities and asked respondents to tick all "forms of locally provided support needed by academic data users." Those who responded to this question (162 or 79 percent of total) reinforced the need for a number of forms of locally provided support, above all "Data discovery / locating sources" (66 percent). All of the answers shown in Chart 5 received "votes" from between one-third and two-thirds of those responding. The average number of needs ticked was three.

An open-ended follow-up question tended to reinforce the forms of support suggested in the questionnaire, although a significant minority felt that no additional support was needed, or expressed concern about where the resources would come from.
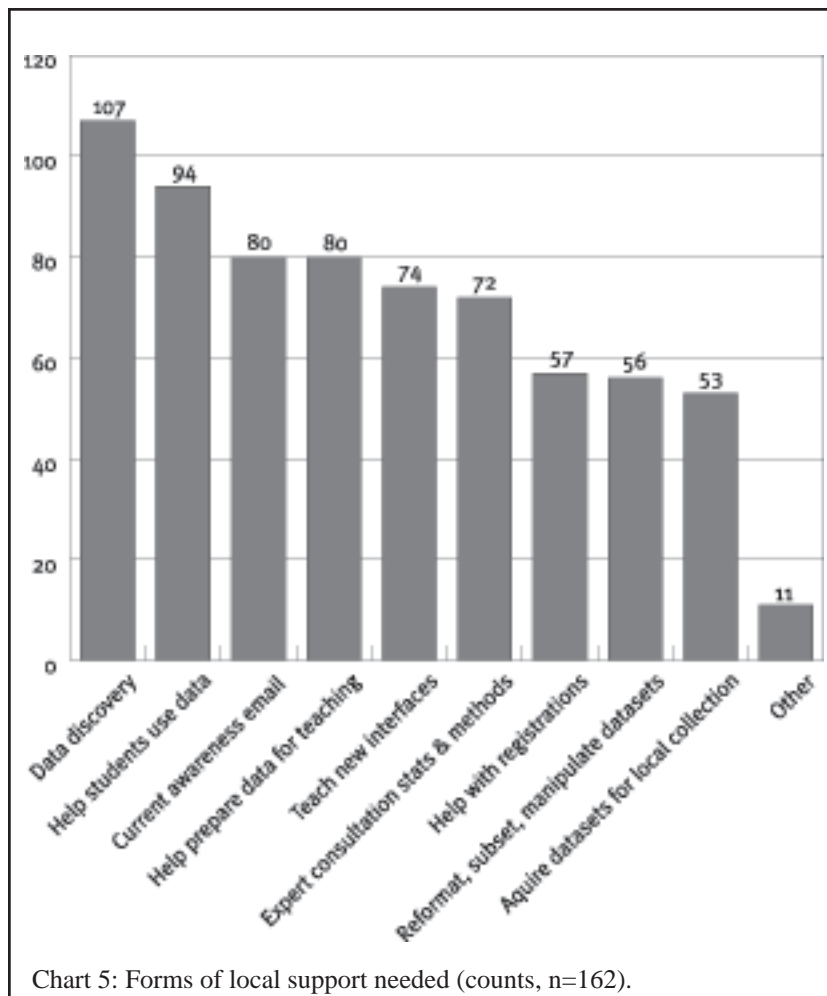
Recommendations

The Task Force and the project team provided the following recommendations to the JISC (project funder) at the close of the project. Further elaboration may be found on the project Web site.

1.   A broad initiative is recommended to promote subject-based statistical literacy for students, coupled with tangible support for academic teaching staff who wish to incorporate empirical data into substantive courses.

2.   The development of high-quality teaching materials for major UK datasets needs to be funded adequately, in order to provide salience to subject matter and demonstrate relevant methods for coursework.

3.   The national data services need to improve the usability of their datasets for learning and teaching.

4.   A more concerted and co-ordinated promotion of the national data services should then follow, which is responsive to user demand.

5.   Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies.

Conclusion

UK higher education is undergoing many changes. The renewed attention to "learning and teaching" is an impetus for change in university teaching practices. Advances in information technology are creating new spaces for learning beyond the traditional classroom, and forms of teaching beyond the traditional lecture. Yet the pressures on academic staff who are still rewarded primarily for research rather than innovative teaching are great. To ensure that statistical literacy is taught effectively, new products and resources must be developed and adequate levels of support and technology provided.

[1] With acknowledgments to the project team: Peter Burnhill (Project Director), Melanie Wright, Sean Townsend; Joan Fairgrieve for statistical analysis; and the Task Force on the Use of Numeric Data in Learning and Teaching. For membership see http://datalib.ed.ac.uk/projects/datateach/participants.html

[2] M. Schield (1999). "Statistical literacy: Thinking critically about statistics.""*Of Significance* (Journal of the Association of Public Data Users):1. Available as of 14 Sep. 01: http://www.augsburg.edu/ppages/~schield/MiloPapers/984StatisticalLiteracy6.pdf

* Paper presented at the IASSIST/IFDO Conference 2001, AmsterdamRobin Rice, Edinburgh University Data Library

Chart 5: Forms of local support needed (counts, n=162).