

An enquiry into the use of numeric data in learning and teaching in UKHE

Robin Rice and Joan Fairgrieve

Summary

An academic Task Force on the Use of Numeric Data in Learning and Teaching has issued a report on the barriers faced by teachers and students to using national data services across a number of disciplines, including but not limited to the social sciences. As the name suggests, the enquiry focused on numeric data, which involves more skills to use than many other types of information resources. Results were analysed from a national survey of teaching departments in universities, and seven case studies of real-life teaching scenarios in both post- and undergraduate classes in several disciplines. The task force contributed views from their own significant experience of teaching in academia as well.

The enquiry was funded as part of the JISC (Joint Information Systems Committee) development programme on learning and teaching, and was conducted between February 2000 and September 2001. The unique focus of this project was on the value of introducing statistical data such as area census statistics, sample survey datasets, and economic trend data to the educational experience of students, particularly when students actively take part in analysing the data, and practice drawing realistic conclusions from empirical evidence.

The enquiry found that despite established use of quantitative secondary analysis of national datasets in research, a number of barriers make its use in teaching and students' independent study difficult, and therefore rare. Whilst print tables and graphs are often used by lecturers in teaching empirical subjects, statistical files requiring 'hands-on' computer analysis are not commonly built into the teaching design, except frequently in methods courses. Yet surely the skills associated with the use of numeric data—such as statistical literacy—are needed along with other “transferable skills” such as information literacy, by today's graduates for them to enter the professions or advanced study.

Only one-quarter of survey respondents who said they used data in the classroom had considered using the nationally funded academic data services provided by the Data Archive (at Essex), MIMAS (at Manchester), or EDINA (at Edinburgh) as a source of the data used in their teaching. The survey uncovered a number of barriers experienced by teachers in the use of these services, namely a lack of awareness of relevant materials, lack of sufficient time for preparation, complex registration procedures, and problems with the delivery and format of the datasets available. These problems were elaborated in open-ended comments by respondents and in the case studies of current teaching practice, and informed the recommendations issued by the task force.

A compounding problem is the lack of local support for teachers who would like to incorporate data analysis into substantive courses. A majority of the survey respondents said that the level of support for data use in their own institutions was ad-hoc. Peer support was more common than support from librarians and computing service staff, and over one-third received no support whatever. The top three forms of local support needed were data discovery/ locating sources, helping students use data, and expert consultation for statistics and methods (for staff).

About the Project

The project was led by Edinburgh University Data Library. Project partners were from three national data centres, EDINA, MIMAS, and the Data Archive, and another university data library at the London School of Economics. The task force, chaired by Prof. Peter Elias of Warwick University, was made up of experienced academics from across the UK who were recruited as volunteers to guide the enquiry and its outcomes.

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. 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.

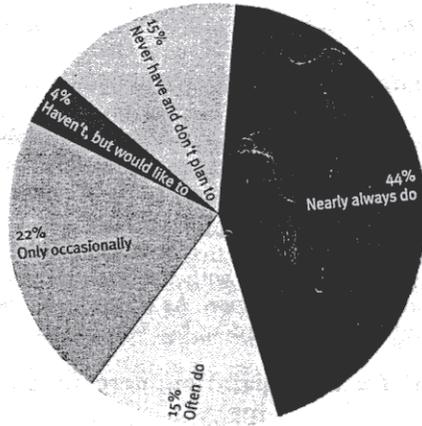
Survey Methodology

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 online 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.

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

Among those who used numeric data in teaching in some form, about two-thirds expected students to work with data 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.

Table 1: Whether course is “hands-on,” by course type.

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

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.

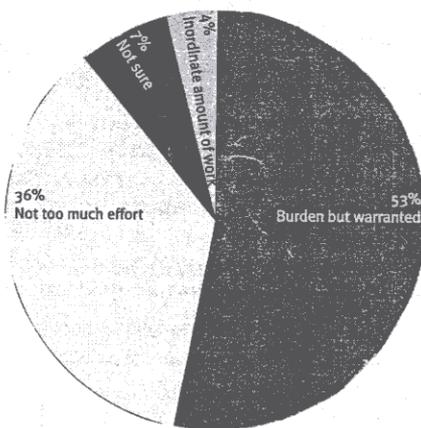
- *Nearly always do (35 percent):* “Statements made need to be backed up with evidence – often of an empirical nature.”

- *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).



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.

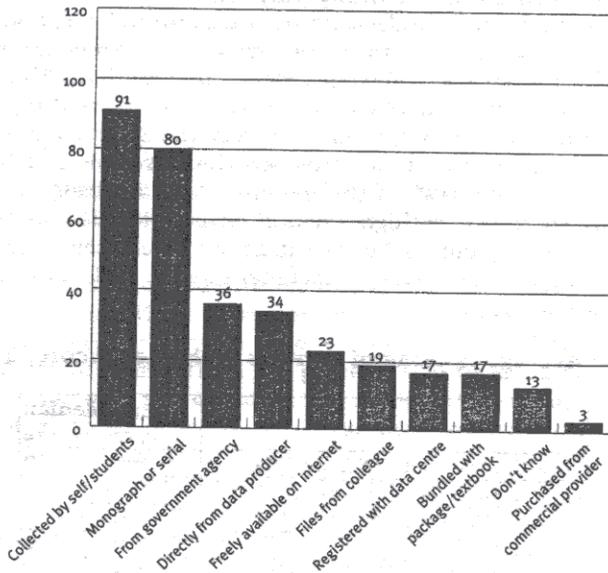
Data Sources and Use of National Data Services

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 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 each 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.

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



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.

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.

Table 2: Average ranking of barriers (8=highest score, 1=lowest).

	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

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.”

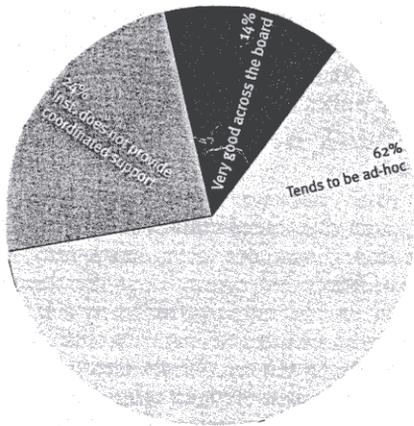
Support Issues

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, computing service, or elsewhere in an institution, but it was not known how much support they actually provide to users.

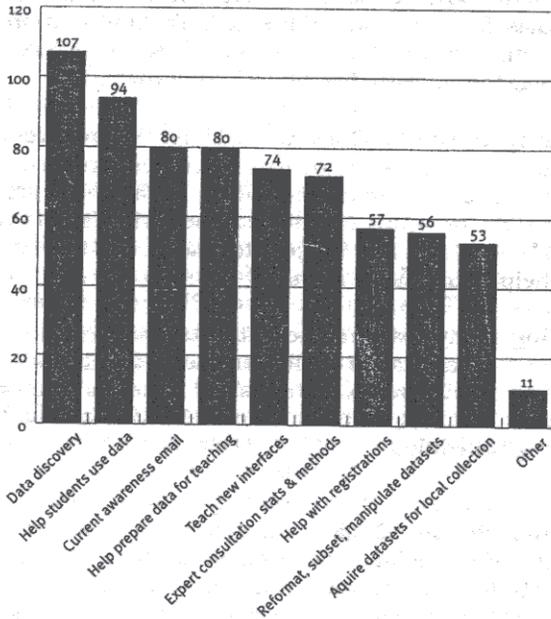
To provide a baseline measure on this issue, the survey asked each respondent, “From whom have you ever had support in obtaining or using data, *whether for teaching or for research?*” Of those who responded, more than a third (37 percent) had received no support at all. More than one source could be 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 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.”

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



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.

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

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 following recommendations were included in the final report, which are addressed to JISC, the national data services, universities, and teaching departments.

- 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.**

- **Key Skills** Responsibility for building students' 'transferable skills' which include statistical literacy, numeracy, critical thinking, data analysis, and computing skills needs to be

addressed right across higher education. Such quantitative-based skills should be integrated with, and not overlooked, in the push for information literacy, IT (information technology) skills, and other skills deemed necessary as educational outcomes, along with discipline-based knowledge.

- **Teaching Rewards** More rewards for innovative teaching are needed, combined with adequate facilities, preparation time, and personal support for teachers who wish to integrate hands-on use of data by students into coursework. This is true for both methods teachers and subject teachers, but the latter may need extra help in making a start.
 - **Training** Many teachers need to build or rebuild confidence in their own quantitative skills for incorporating students' use of data into coursework. 'Refresher' courses should be made available locally, which are convenient for staff with busy teaching schedules.
 - **Student-centred** Both undergraduate and postgraduate students should be given adequate support for the use and analysis of secondary data sources as part of their independent research. It is unrealistic for support for students to fall solely on the tutor or supervisor, because there are usually several learning curves that need to be mastered by the student in order to get the empirical result desired. (A majority of survey respondents desired help at the local level with both data discovery / locating sources and helping students to use data for learning and research.) This would help lead to a more student-centred education as well as reduce the burden of teachers.
2. **The development of high-quality teaching materials for major UK datasets must be funded adequately, in order to provide salience to subject matter and demonstrate relevant methods for coursework.**
- **Fully-documented datasets** Data-related teaching materials need to include subsets of large complex datasets, along with clear documentation about the original and subsetted dataset, practical exercises for students, and teachers' notes. If teachers are not confident about what the dataset can demonstrate to the class, they will not use it. They also need to be in a usable format for the local environment.

- **Reinforce taught subjects** Subsets need to be tailored to a range of subject disciplines, able to illustrate concepts that are actually taught. The differences in needs and purposes between methods-courses and subject-based courses found in the survey need to be taken into account when designing teaching materials, and the greater reluctance of subject-based teachers to incorporate hands-on work with data.
 - **Supply interesting evidence** Quantitative study has a reputation for being dry, irrelevant, or even dishonest among many students ('Statistics lie'). A related problem is that at present, easy-to-use sample data available in standard statistical packages are either US-based or outdated. This can be combated through provision of current, interesting data, based in the UK or other geographic area of interest to the students. If enough user-friendly subsets of major studies are developed for learning and teaching, students can be encouraged to use more empirical data in their own research without added burden to teachers, leading to a more learner-centred education.
 - **Define responsibility** It is unclear at present who should lead the effort to create these new data-related teaching materials, but national data services have expertise regarding datasets they service, while LTSN (Learning and Teaching Subject Network) subject centres have knowledge of teachers' needs in particular disciplines. Teachers themselves may have much to offer, given sufficient resource for development and an environment which encourages sharing. Learning and teaching materials should be free at the point of use to encourage uptake, but funding needs to be earmarked within existing structures for materials to be developed.
 - **Develop partnerships** One possible model for the creation of new teaching materials is currently being used by some of the other projects funded in this programme (see Appendix A). Partnerships between data centres and subject centres have been established to commission teachers or learning technologists in universities to develop and contribute on-line, locally customisable course materials for shared use. Other members of the academic community will then be encouraged to share their customised learning units, and newly authored units, with the rest of the community.
3. **The national data services need to improve the usability of their datasets for learning and teaching.**

- **On-line tools** Intuitive search and extraction interfaces to downloadable data need to be developed that comply with current expectations of accessibility. At present, web-based access meets that norm. The Data Archive, MIMAS, and EDINA could also look for partnerships that pilot students' use of data within new technologies such as VLEs--Virtual Learning Environments. (This could provide exciting examples for teachers and learning technologists working together to modernise courses.)
- **Customise** Datasets need to be provided in a variety of useful formats, conforming with software that is supported in local environments and appropriate for the course of study.
- **Reduce time-lag** Delivery time for user access needs to be improved to be realistic for the time pressures of students and teaching staff.
- **Streamline student registration** Registration procedures should be more user-friendly (currently they are depositor-friendly). Terms and conditions of use should be simplified for learning and teaching purposes, eliminating individual student usernames and signatures when publication of work is not an outcome (i.e. letting the teacher sign on behalf of a class, taking responsibility for usage).
- **Profile without paper** Technology should be favoured for monitoring usage over giving users more paperwork to fill out. Registration must be recognised as a barrier to use in learning and teaching, stripped down and eliminated whenever possible. The current model of providing access free at the point of use however, should be kept.
- **Train the trainers** Site representatives or other relevant support staff should be offered training in use of the national data services that is commensurate with their own professional development goals and their department's policies (e.g. travel, days away, and course fees are potential hurdles.) These 'champions' can then offer group or individual support to academic staff and students at the time of need.
- **Inter-operability** Cross-references and web links should be used between data providers and data discovery 'portals' to increase awareness of relevant resources. New technologies and

standards should be exploited that can generate links between numeric data sources and to and from text-based sources.

4. **A more concerted and co-ordinated promotion of the national data services could then follow, which is responsive to user demand.**
 - **Effective marketing** Marketing is not just finding a better way to tell people about your service. The Chartered Institute of Marketing defines it as ‘a management process which identifies, anticipates and satisfies customer requirements profitably’ (or efficiently). This involves understanding users and their motivation, segmenting them into groups with common needs, and developing strategies with clear objectives.
 - **Roadshows** Regional roadshows can provide stronger links with local support staff and accessible opportunities for overburdened teaching staff. By providing a hands-on component to the promotional event, the national services can offer users a low-investment learning opportunity along with their chance to ‘stump’ to an audience.
 - **Local templates** National data resources need to be ‘localised’ to improve their accessibility within the education environment. For example, catalogue records of important datasets could be provided to university libraries for inclusion—whether or not there is a local dataset collection or ‘data library’. Promotional materials should be customisable at the local level to enhance targeting of different user groups. The emphasis should be on adding value to local awareness-raising efforts rather than national or organisational branding.
 - **Dialogue** Consensus needs to be sought among national and local ‘gatekeepers’ and stakeholders, about appropriate publicity and promotion by different organisations within the DNER (Distributed National Electronic Resource) and beyond, to avoid confusion as promotional efforts are stepped up. Dialogue with research councils, national and local government agencies, scholarly and professional societies is also needed.
5. **Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies.**

- **Identify locus of support** Universities need to identify appropriate staff for data and statistical support for both research and L&T. Data-related support need not be centralised in an academic data library, but it needs to be easily identifiable by users. Sufficient resource needs to be allocated for data-related support functions whether they are located in libraries, computing services, or specialist departments. Support staff in both libraries and computing services should be encouraged to 'keep up' with national developments in order to communicate relevant services to users at the point of need, as a component of professional development policies.
- **Balance resources** Increased resource is being found for development of managed learning environments (MLEs) and Virtual Learning Environments (VLEs) in institutions to keep apace with student and staff expectations in learning technologies. Along with other concerns, the ease with which numeric data and other empirical evidence can be used in these electronic environments needs to be considered. Innovators should be encouraged to present examples of successful modes of data presentation and analysis in these new learning environments.
- **Support for students** Whilst data-related support for academic staff may be slipping through the cracks, in many institutions support for students is non-existent, particularly undergraduates. For teaching departments to construct ambitious learner-centred curriculae, they need to know their students will have access to computing resources and be able to get personal support beyond their own office hours. Dialogue needs to take place across university academic and support departments to determine if support levels are adequate for all and to identify solutions for new learning goals and environments.

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.

Robin Rice
Data Librarian
Edinburgh University Data Library
Main Library Bldg.,
George Square
Edinburgh
EH8 9LJ
Tel 0131 651 1431
R.Rice@ed.ac.uk
<http://datalib.ed.ac.uk>