Survey of the State of Analytics in UK Higher and Further Education Institutions 2013

Introduction

An informal survey was undertaken by Cetis in May and June 2013. Subscribers to a number of email circulation lists - with members coming largely from institutional IT, administration and educational technology responsibilities - were invited to respond.

The purpose of the survey was to:

- Assess the current state of analytics in UK FE/HE.
- Identify the challenges and barriers to using analytics.

For the purpose of the survey, we defined our use of "analytics" to be the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data. In practical terms, it involves trying to find out things about an organisation, its products services and operations, to help inform decisions about what to do next.

Various domains of decision-making are encompassed: the kinds of decision that is readily understood by a business-person, whatever their line of business; questions of an essentially educational character; and decisions relating to the management of research. The line of questioning was inclusive of these three perspectives. The questions asked were:

1. Which education sector do you work in (or with)?
2. What is your role in your institution?
3. In your institution which department(s) are leading institutional analytics activities and services?
4. In your institution, how aware are staff about recent developments in analytics?
5. Do the following roles use the results of statistical analysis such as correlation or significance testing rather than simple reporting of data in charts or tables?
6. Which of the following sources are used to supply data for analytics activities?
7. Which of the following data collection and analytics technologies are in place in your institution?
8. Please name the supplier/product of the principle software in use (e.g. IBM Cognos, SPSS, Tableau, Excel)
9. Which of the following staff capabilities are in place in your institution?
10a. What are the drivers for taking analytics based approaches in your institution?
10b. What are the current barriers for using of analytics in your institution?

The survey is informal in that:

- no attempt was made to have a balanced sample, either in terms of institutional type or respondent role;
- it relied on voluntary participation, so will suffer from selection bias;
- multiple responses from a single institution may have occurred, but these cannot be identified;
- we surveyed the knowledge of individuals rather than the actual state of an organisation.

These facts, coupled with the small number of responses, means that the resulting data cannot be assumed to represent the true state of affairs but to be indicative. The report is written as a stimulus both for discussion and for more thorough investigation into some of the areas where the survey responses hint at an issue.

Terminology: a "response" is a single submission of the survey and an "item response" pertains to a single question.

Basic Facts

There were 26 responses.

Which education sector do you work in (or with)?


What us your role in your institution?
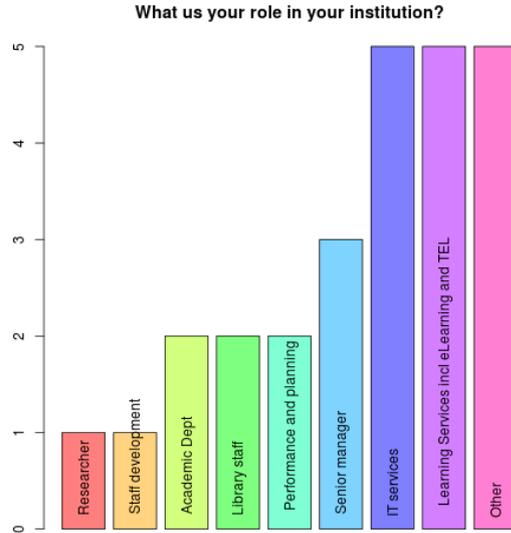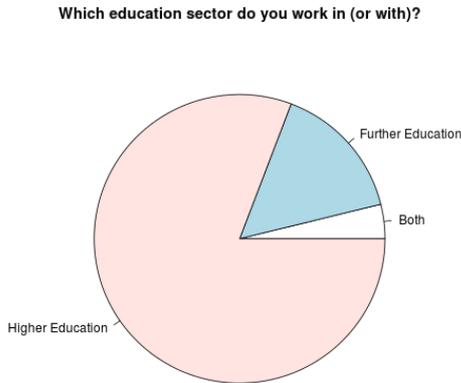
Concerning the State of Knowledge

The line of questioning sought the knowledge of the respondent and not the actual state of affairs in a given institution. Although this fact limits the extent to which responses can be interpreted in relation to the question as asked, it does allow us to form a picture of the way knowledge of analytics activities is distributed. We can form a response to the question: what does "don't know" look like?

Technical note: most questions contain multiple parts and for these a "don't know rating" is calculated as the proportion of parts within that question that were answered "I don't know". Question 3 contained no options so has a don't know rating of 0 or 1 only.

The mean don't know rating for each response was calculated and found to have a median value of 0.313 indicating that half of the sample had more than around 31% "don't know" item responses. 25% of the responses had around 17% or less "don't know" item responses. The conclusion is clearly that there is a lot of uncertainty. This is particularly striking because the respondents were self-selecting; it would be expected the people with the least knowledge would be less likely to respond.
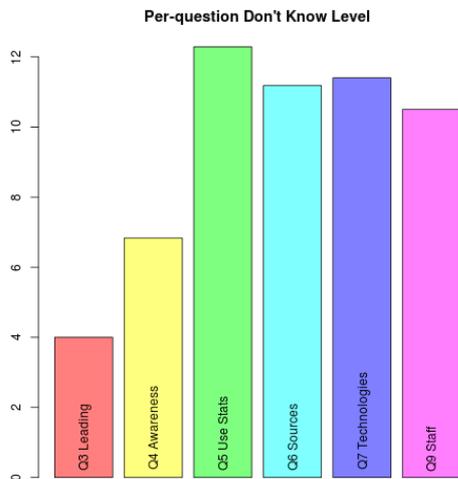

The state of unknowing varies between questions. This plot shows the mean over all responses for each question.

It seems to indicate that there is a better awareness that there is something going on, both within the institution and generally (questions 3 and 4) than there is about the way an institutional response is being implemented (questions 5, 6, 7, 9). A more detailed analysis appears to indicate that questions 5 and 6 are a particular hot-spot of uncertainty, with a probable cluster of responses where the item responses for both questions were "don't know" whereas other items responses where known. In many cases, uncertainty is spread over many questions without any pattern being evident.

Leadership of Analytics Activities

In your institution which department(s) are leading institutional analytics activities and services? (multiple selection was possible)
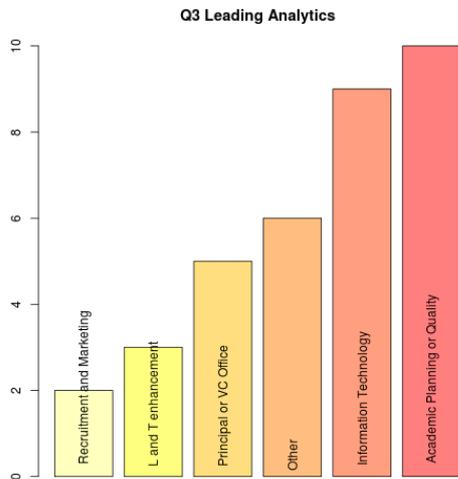

The dominant leaders reflect the centres for key institutional data and the IT services to handle it. There were only 4 "don't know" responses and the "other" item responses show some partial overlap with the specified categories: "Central department", "Academic planning or quality (APQ), Student Information (Registry)", "Schools leading learning analytics, IT leading institutional analytics", "Teacher Training Area", "Compliance Unit, Learning Services", "Pockets within certain projects".

Awareness

In your institution, how aware are staff about recent developments in analytics? (6 sub-questions, according to staff role)

An impression of the general level of awareness is given by summation over all staff roles given in the question: (Executive Management, Academic Managers, IT Service Staff, Teaching Staff, Library Staff, Research management).
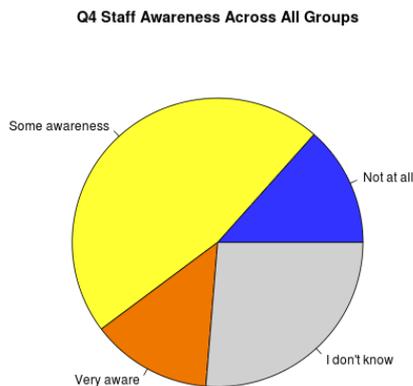
A stacked bar plot, ordered by increasing number of "Very aware" responses, shows the different levels of perceived awareness within each staff group:
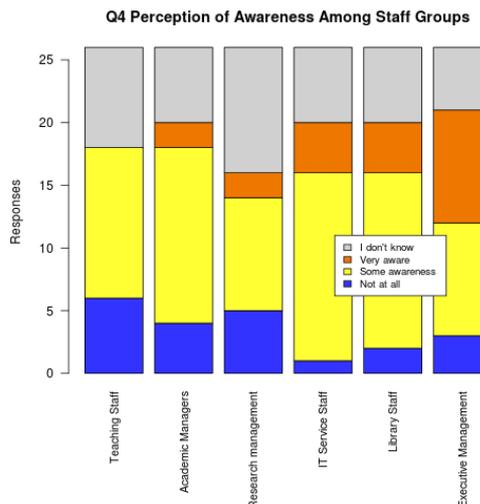
These results are likely to be biased because of the distribution of backgrounds of respondents and this may mean that we should not assume the item responses that were not "don't know" are representative.

The tentative conclusion seems to be that, unlike many initiatives with a strong data and IT element, executive management awareness is strong. In general, the management and support roles appear to be more aware than the academic staff but the similarity between IT and library staff may not have been expected. This data suggests anyone planning to develop analytics at an institutional or sectorial level should not make too many assumptions about awareness, rather they should investigate the state.

Use of Statistical Analysis

Do the following roles use the results of statistical analysis such as correlation or significance testing rather than simple reporting of data in charts or tables? (7 sub-questions. according to staff role)

Summation over all staff groups (Executive management, Academic management, IT services staff, Teaching staff, Library staff, Facilities and estates staff, Finance and purchasing) allows us to capture the overall impression.


Q5 Use of Statistical Analysis Across All Groups
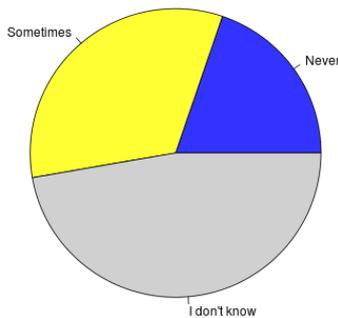
The general levels of "don't know" are really rather high. What does this tell us? Probably that reports and verbal communications do not refer to this kind of analysis. Whether this means that it is not occurring or that it is being done and not communicated would require further investigation. Arguably, whichever is the case we should be concerned; this kind of analysis is often necessary to decide whether the tables of data and charts really signify a compelling case for action. Furthermore, wider use of these well-established and routine statistical methods within an institution should probably precede use of more advanced predictive methods.

The following plot drills down into the roles and shows that the pattern is similar across several roles but has facilities/estates and finance/purchasing at an extreme. Arguably, these areas should be making more use of statistical methods than the data indicates since these areas are the ones where this is most easy to do without the complexity/subtlety of teaching and learning or the uncertainties in external strategic factors; this may be an artefact of the respondent population simply not knowing what goes on in these functional areas.

Q5 Use of Statistical Analysis Among Staff Groups
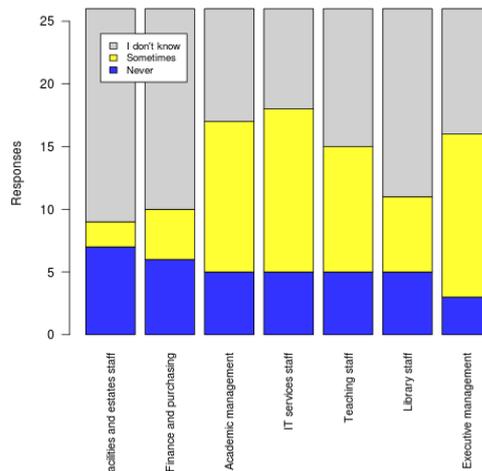
Sources of Data

Which of the following sources are used to supply data for analytics activities? (11 sub-questions, according to data source)

Again, we must be cautious about interpreting the results; there is a large "don't know" fraction and the distribution of respondent roles/departments will have introduced bias that cannot reasonably be estimated.

The data is ordered according to the number of "yes" responses for each sub-question and shown in the following chart. Sector data includes data from the national student survey, HESA, UCAS etc.


Q6 Which Data Sources are Used?
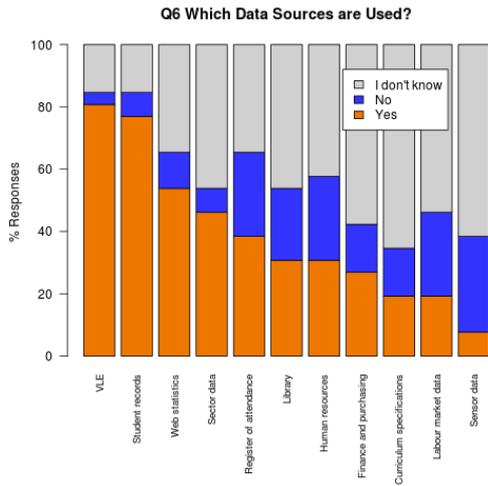
Although we should be cautious in assuming anything about the actual situation that the "don't know" responses mask - for example that the don't know fraction can be assumed to be "no" or that it can be assumed to have the same proportion of "yes" and "no" as these definitive responses - this plot suggests a few provisional comments:

- VLE data is widely used yet it only gives a partial and ambiguous view on learner activity. Given the apparent lack of use of even basic statistical methods, are we being over-confident about the value of this data?
- Collectively, VLE, attendance, and library data would give a more balanced picture of engagement yet the second two are much less widely used.
- Given the availability of sector data, what is the explanation that only half of the responses indicated it is used?

The apparently low use of labour market data may be a missed opportunity.

Technologies in Place

Which of the following data collection and analytics technologies are in place in your institution? (5 sub-questions, according to technology)

Q7 Which Technologies are in Place?
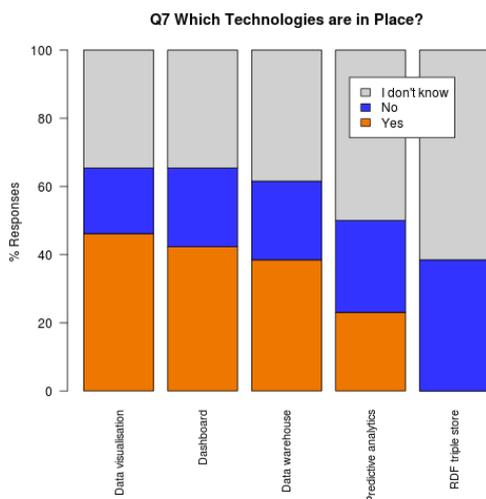
There is some consistency across the more common technologies given in the question: data visualisation, dashboards and data warehouses. The less commonly used technologies are those with the largest number of "don't know" responses although it is plausible that some "don't know" responses for "predictive analytics" occurred because the term is imprecise. As before, we should not assume that the relative proportion of yes/no responses gives an approximate estimate of the proportion of institutions that actually employ these technologies. For data visualisation and dashboards, it might be reasonable to assume that many "don't know" responses should be substituted with "no", given that these are user-facing manifestations of analytics.

Survey respondents were asked to provide information about products/suppliers as textual item responses. Each response could contain any number of product names. The following plot shows the number of item responses for products that were named two or more times. More than 1 supplier may have been specified in a response.

SPSS is now part of the IBM Cognos offering, although it may be used independently of the suite.

Q7 Technology Products/Suppliers
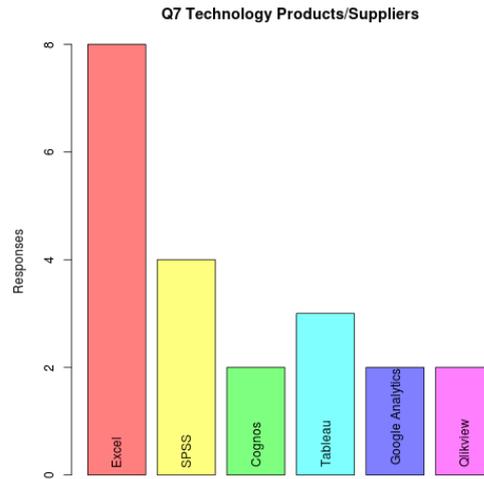
Drivers for Analytics

What are the drivers for taking analytics based approaches in your institution? (multiple selection was possible)

Q10 Drivers for Analytics
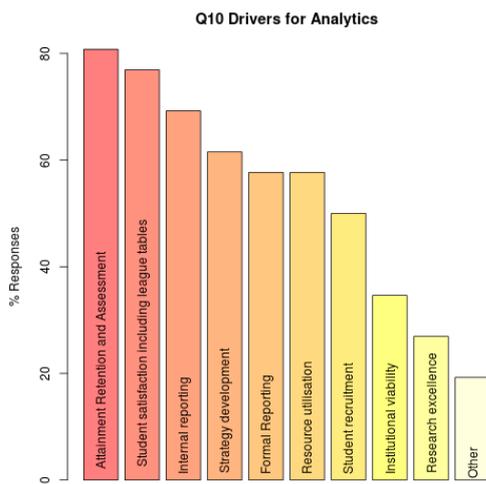
The survey responses match the topics typically discussed in the literature on analytics in universities and colleges. The following suggested drivers attracted lower levels of positive response than they may deserve:

- The low value for research excellence may reflect the survey group, which contains few participants with research management responsibility and an unknown number of responses from research intensive institutions.
- Student recruitment seems to be a missed opportunity. The 50% response rate for this driver should be verified.

Resource utilisation (excluding "human resource") is another candidate for verification. If the <60% rate is correct, it suggests opportunities for efficiency savings are being missed in too many institutions.

Barriers to Adoption

What are the current barriers for using of analytics in your institution? (multiple selection was possible)
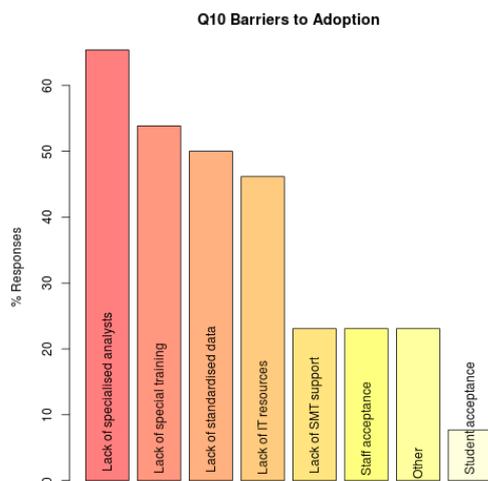
Bearing in mind that survey respondents were self-selecting, and so may have been disproportionately positive in their attitude towards analytics, the results seem to indicate that several common barriers to IT and data based initiatives arising from the attitude of people are not prevalent. The issue of lack of senior management report, which is often seen as a significant challenge, is not a major factor. This matches the earlier-noted level of awareness that is perceived among executive management. Staff and student acceptance is also not seen as an issue. This may be due to there being little practical effect on day-to-day teaching, learning, research, and support because few analytics initiatives have been rolled-out at institutional level. Even if we assume the survey is accurate, it is plausible that attitudes could change dramatically.

Whereas the attitude of people is not generally perceived as a barrier, both IT and analytical capabilities are. Among the respondents, there is a slightly greater sense that the analytical capabilities - specialised analysts and training - are common barriers than IT-related aspects - standardised data and IT resources.

A more detailed inspection shows some evidence that the IT and analytical barriers occur together whereas the attitude-based barriers are apparently randomly distributed. Three respondents only identified specialised analysis and training as barriers. No response identified zero barriers.

Respondents who stated "other" supplied: "time", "enough staff resources and priority to drive forward", "No idea, as we aren't there yet", and three counts where the field was left empty.

Supplementary Information

This report was produced by Adam Cooper and the survey was developed by Li Yuan and Stephen Powell. All are members of Cetis and the work was supported by Jisc. It is licensed using the Creative Commons Attribution Licence: http://creativecommons.org/licenses/by/3.0/.

Survey Form, Data and Source Code

These are all available from GitHub (https://github.com/arc12/Cetis-Analytics-Survey-2013):

- Survey form (PDF).
- Raw data (CSV).
- Source code for R[1] with the cluster package[2]. This is intended for processing the raw data using knitr, which is conveniently done using RStudio (http://www.rstudio.com).
- A slightly extended version of this report containing some additional exploratory plots.

[1]: R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
[2]: Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2013). cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4