ALT-MEMBER Responses to Jisc Private Consultation on Code of Practice for Learning Analytics

This is feedback for Jisc's private consultation on 'Code of Practice for Learning Analytics v0.4' from ALT Members collected via the ALT Members discussion list and collated by Martin Hawksey as Chief Innovation, Technology and Community Officer. As well as this private response to this consultation ALT will also revise and publish as part of the public debate. In addition to this feedback we'd also like to highlight the importance of Learning Analytics that members expressed in the 2014 ALT Annual Survey, see http://repository.alt.ac.uk/2358.

Key points

1. **Scope of the Code of Practice:** we refer back to ALT's definition of Learning Technology as the broad range of communication, information and related technologies that can be used to support learning, teaching, and assessment. ALT would like to see the Code of Practice for Learning Analytics to reflect a similarly broad understanding of what is meant by Learning Analytics including teaching and assessment, with consideration for more varied role holders, as is highlighted in the general feedback below;
2. **Ownership of analytics and data**: inclusion of statements around ownership and access to data and analytics within the draft are a very welcome. These are currently mostly of the benefit of the institution and should be more balanced between the institution and the student. The relationship between third parties and vendors and the institution should be made more explicit;
3. **Exemplars of practice:** as an emerging field definitions and methods may not have common understanding. As such the Code would benefit with the inclusion of exemplars to illustrate good/bad practice or link to existing resources with illustrate these. Examples may also be drawn from non-Jisc resources such as outputs from the LACE Project and the Open University's Principles for the Ethical use of Student Data.

Full narrative feedback

General

A concern raised is that the Code of Practice for Learning Analytics assumes that everyone knows what they are doing with analytics, which they do not, and that there are actually appropriately qualified staff and students to review changes.

The draft code also does not seem to take into account staff metrics/data and is very student focussed. It presumes people will be doing "learning analytics" and not "teaching analytics" and I think it is the latter that most people are actually doing.

Whilst the current code addresses highlighting the measures to make analytics valid it is also suggested the code references common pitfalls such as sample sizes, strong halo/placebo effects and reliability of predictive techniques.

Whilst there is existing legislation covering data protection the current draft code does not make any specific reference to agreements held with the host institution and any third parties. In particular the code should make explicit the protection of student data when using intermediaries for data storage or 'analytics as a service'. As part of the code it would be useful for institutions to make it clear how student data is used by third parties, for example, services like Turnitin use data from all submissions. The development of a separate vendor code of practice which institutions can ask suppliers to conform to would also be a very welcome development.

For all analytics/process/algorithms etc, there is a need to convey a shared purpose with the student, preferably one that centralises their progress in learning-be nice to see that spelled out a little bit more in this charter. Student representatives appear to have input, but **exemplars** as to how processes might be carried out and why, might demonstrate how transparency and in turn how validity can be achieved in terms of all stakeholders.

Is it also worth outlining what is meant by Learning Analytics/Formative/Summative assessment, and those demarcation points, in terms of

  a. learner progress
  b. aggregation for analysis

In view of: Learning Analytics used for formative rather than summative assessment. This thinking stems from involvement in staff/student consultative committees. So, when the process of Learning Analytics is used for course enhancement, for example, it is always nice to be able to draw on associated charters and explain to the student rep what processes are in place and why. If there are clear exemplars on charters this can highlight processes that validate what the institution does too.

Given the number of Jisc Student Innovation project proposals to develop student-owned interfaces on organisationally-owned data, it suggests an appetite among digitally-savvy students for the actual data to be exposed to them, and not just via institutional dashboards and interfaces. This facet is something the code might like to make more explicit.

Transparency and Consent

> *Where possible institutions should also make the algorithms transparent and/or clearly describe the processes involved in producing the analytics.*

If the meaning of 'transparent' includes "and the **implications** of these algorithms are evident", then this is sufficient, but if it simply means (or could be interpreted as meaning) that the nature of the data stored and algorithms performed is open, then this is not really sufficient.

Similar issues have been experienced in HCI and includes examples where anonymisation and aggregation, which are normally seen as 'good; things to do for privacy, can actually increase the personal sensitivity of data[1]. Also cases where the use of machine learning or neural networks can potentially lead to situations where decisions are illegal (e.g. gender/racial discrimination), not to mention unethical[2].

It would be useful to include some wording that puts a reasonable level of onus on the institution to seek

---

[1] http://alandix.com/academic/papers/int90/
[2] http://alandix.com/academic/papers/neuro92/neuro92.html

to understand and share the potential **consequences** of algorithms.

It was commented that "where possible" is a get-out clause, transparency must be a priority, if we cannot be transparent about how and why, then we should not collect the data.

Privacy

> *The use of "sensitive data" such as religious affiliation and ethnicity for learning analytics will be restricted and for clearly specified purposes.*

It is difficult to see how this is an exemplar for specified purposes, when such variables are not normally correlated with learning outcomes, in order to be shared with the learner to enhance progress? Thus, it is difficult to understand what is meant by validity in this document. Perhaps better examples might highlight what is meant.

Access

> *Students should be able to access all learning analytics performed on their data in meaningful, accessible formats, and to obtain copies of this data in a portable digital format. Students must have the ability to correct inaccurate data held about themselves.*

> *They should normally also be able to view the metrics and labels attached to them. If an institution considers that the analytics may have a harmful impact on the student's academic progress it may withhold the analytics from the student, subject to clearly defined and explained policies.*

The drift of the first paragraph is very much to be welcomed. Providing learners with access to their own analytics data can, if done well in terms of usability, be a powerful tool to help them reflect on their own learning -- and thereby learn about learning. As standards evolve, there may also be opportunities for learners to maintain their data in Personal Data Stores and use their learning profiles to help configure/optimise other learning experiences from other providers/publishers.

The second para remains too weak on the protection of learners' position, and too skewed towards institutions either being too lazy or too defensive to provide access to learners' personal data. It makes the assumption that the institution knows better than the learner what is in the learner's interests - this is a questionable assumption and dangerous in that it is open to abuse. The assumption should be that all metrics, labels and other data about learners should be available to them by default. The onus should be on the institution to make a case for going against this. They should be required to do this *in* terms *of policies that are defined by some entity (Jisc?) other than the institution itself*. Institutions should not define their own laws as well as being judge, jury and defence barrister.

It was also commented that rather than withholding information, why not insist that wellbeing structures are in place to ensure that pastoral processes are in place.

Action

The section on 'Action' could be strengthened by focusing clearly on positive uses. As it comes before 'Minimising adverse impacts' it could be called something like 'Enabling positive interventions'. It could set out clearly what are legitimate uses of student data and why they are beneficial. Main uses have already been identified as:

1. Helping individual students based on their own data
2. Merging individuals' data with those of others to help the group
3. Using data to help future cohorts of students

Each of those could be expanded a little to explain what can actually be done with data and why it is beneficial that it should be done. This will then illuminate the need for another section, which you put under the heading 'Human intermediation' in the blog post[3]. These issues are subtle and might be difficult to enshrine into a code, but it seems important to say that the data system itself - however ethically and efficiently managed - does not benefit the learner. Benefits are provided when real learning support is offered, or guidance given, to individual students who are identified as being in need. Also when teachers adapt their teaching responsively to meet the identified needs of a cohort, and when curriculum teams adapt the curriculum offer to meet the changing needs of the student body as a whole. And again, of course, when the learning environment is adapted in response to identified needs. All of those things demand human expertise and time, which have to be invested in. So while it might be a bridge too far for the Code to say 'data analytics is not student support on the cheap', it could say something about the need for staff to be able to interpret and use the outcomes of analytical reports to benefit students individually and collectively, and to have the time, resources, support and expertise to do that.

Stewardship of data

> *Data for learning analytics will comply with existing institutional data policies and the Data Protection Act … students should have the "right to be forgotten"*

As well as making reference to EU legislation on the 'right to be forgotten' it would be useful to highlight other parts of legislation and implications for learning analytics. In particular the practice of 'profiling' appears to have direct implications with Learning Analytics[4].

It was also commented that what recommendations are there as part of the code for monitoring stewardship of data.

Document Information

Version	Date	Author	Notes
1.0	2015-05-05	Martin Hawksey + ALT Members	

This document is licensed for reuse under a Creative Commons Attribution-ShareAlike 4.0 Unported (CC BY-SA 4.0) licence. You are welcome to adapt it for your own purposes, in which case please acknowledge this document as your source.

---

[3] http://analytics.jiscinvolve.org/wp/2015/01/30/developing-the-code-of-practice/
[4] http://www.out-law.com/en/articles/2015/february/eu-data-protection-reforms-the-implications-for-profiling/