

Data Collection

This project has been designed to use data relevant to performance and engagement collected while the student is at university. It does not seek to explore data or performance which fall outside of this objective such as ethnicity, gender, socio-economic status or high school performance. If you wish to use these measures we recommend commencing a separate project and keeping it distinct from the early warning system unless you are familiar with R and model building procedures.

Step 1: Identify historical data and their predictive power

Identify what historical data has been recorded and is available to use. Some examples of data sources which produce useful predictors can be found in Table [1.1]. Contact administration and support staff to find out what information they record and have available and what IT can provide for previous years. This information will be used to assess the predictive power of these measures before live implementation – see [calculating predictive power].

Step 2: Identify what data is currently available

Identify what information could easily be obtained for the upcoming year or measures which might be useful but are not currently collected. Admin staff can provide a wealth of information in this area. Ask about things they often find themselves chasing students for but don't officially record. Speak to the IT department to find out what information they can provide from your VLE software.

What kind of information are you looking for?

When searching for this information you will likely find more data than you can use, particularly VLE data. To make life easier for you and your colleagues when requesting historical, current or future information ask colleagues to consider the format and how they are providing data. Your data will ideally be:

- Simple** – giving you data in its simplest form will save you the time and effort of having to dredge information to identify the useful parts. For example, total number of absences is more useful than a comprehensive list of all absences.
- Comprehensible** – You may need to request a specific format or a full explanation of what you are receiving. For example, IT systems often export raw data which means that column headings are based on the system code, which can be difficult to decipher, rather than the alternative full text which is displayed in the VLE. Similarly, administrators may head columns with titles they understand but are undiscernible to others.
- Accessible** – preferably (eventually!) all your data will be stored centrally or you know who can get the information to avoid the requirement of multiple team members obtaining information from multiple sources. Making exports from VLE customisable also saves time.
- Discriminatory** – your data should discriminate between students. For example, nominal data describing completion/non-completion of logging in to the VLE is not useful as almost all students will do this. In this case ratio data would do better at discriminating. Rather than ask 'Has the student logged in to the VLE system?' ask, 'What is the time gap between induction to the VLE and first log in?'

The following table provides examples of sources of information to be used as predictors across four categories. Ideally, obtain at least one from each of the categories to help identify academic difficulties, pastoral issues and motivational or attitude issues all of which can contribute to disengagement and academic failure.

Category	Description	Sub-category	Examples
Mandatory Activities	Why? It is known that poor performance in completing mandatory tasks predicts academic failure risk[1].	Administrative	<ul style="list-style-type: none"> • Ethics forms • Plagiarism forms • CVs • Collection of ID cards • Photo submissions
	What? Any task essential to student progression which does not give a mark. Mandatory tasks have high uptake so try to collect ratio data where possible. E.g. rather than record submission/non-submission, record promptness of submission	Academic	<ul style="list-style-type: none"> • Progress exams • Formative exams • Quizzes • Practical tests, e.g. hand-washing • Admission tests
Attendance and Punctuality	<p>Why? Patterns of Absence (A) and Lateness (L) may indicate underlying difficulties</p> <p>What? Any course-related activity where attendance or punctuality can be recorded is suitable here. In particular look for <i>unauthorised</i> (failure to notify) A/L</p> <p>NB: stick to general data rather than recording reasons for absence</p>	-	<ul style="list-style-type: none"> • Tutorials • Seminars • Lectures • Submissions e.g. proposals for projects <p>Look for patterns of A/L in day/individual courses/total numbers</p> <p><i>Authorised</i> – days of advanced warning given</p>
Student engagement	Why? Students are expected to engage but often don't. Lack of engagement may indicate that the student orientates themselves to the minimum requirement and is focused on summative exam learning only.	Staff	<ul style="list-style-type: none"> • Completion of surveys • Attending meetings with personal tutors
	What? Look at how students engage with course content which is not mandatory but expected, e.g. staff-student meetings and VLE usage. NB: Some indicators (VLE particularly) can be misleading. Look for pattern of usage rather than simple frequency/duration indicators	VLE	<ul style="list-style-type: none"> • Promptness of first log in • Frequency of visits • Length of visits • Time of visits • Attendance at online voluntary events

Voluntary activities	<p>Why? Students who undertake activities voluntarily go beyond the expectations of engagement meaning they are generally motivated and proactive. Those who don't are not automatically at risk but these indicators will help differentiate between those most and least at risk.</p>	General	<ul style="list-style-type: none"> • Introduction to the course • Becoming a year representative • Peer support activities, incl. peer families, study groups etc.
	<p>What? Anything not part of the core course requirements (but still relevant). Record ratio data where possible. However, nominal data may still discriminate between least and most at-risk students depending on activity uptake.</p>	Academic	<ul style="list-style-type: none"> • Participation in optional reports, exams or tutorials which are not summatively assessed • Additional projects not part of core course requirements

Table 1.1: Exemplar indicators by category

Making sure data collection goes right

The data you collect will help students. However a large amount 'ordinary' data combined can give an incredibly detailed picture of a student. Therefore you will want to avoid particularly sensitive data, for example:

- IP addresses for location
- Reasons for absence, particularly if they are medical or personal

Knowing too much could create a conflict between your academic role and your duty of care. Think carefully about what you need and what best meets those needs.



Hints and tips for data collection

- Try to be conservative rather than exhaustive in data acquisition – if you aren't sure how much time you can spend on the project start small and trial a simple system first
- Think carefully about what questions you want your data to answer and what would be most and least appropriate
- If in doubt about the suitability of the data, consult with your local data protection officer or ethics committee
- Consider allocating a non-assessing member of staff to collect and manage the data to avoid potential teacher-student relationship issues

Version Changes

Version	Date Released	Changes	Editor
v0.1	May 2015	New document	Avril Dewar