Tiger stripes: what should performance benchmarks look like?

Tom Karmel
Introduction

NVCER has been interested in publishing institutional level statistics for some time, and a number of statistical publications have been released that provide data for individual providers but without names. This limits their usefulness because the whole point of individual provider data is so that readers can make comparisons across institutions. The reason for NCVER's current approach is that the current VET data protocols proscribe the publication of data for individual providers without the permission of the provider. This protocol has meant that, unlike for higher education where tables of indicators have been the norm form many years, that there has been a complete absence of performance indicators for the VET sector.

The Commonwealth government signalled in the 2010 Budget that this situation is about to change, with funding provided for the development of the MySkills website for VET providers.

Performance indicators need attention on three fronts. The first is the range of indicators that should be considered. The second is the unit of analysis - the provider as a whole or individual subject area. The third is the issue of statistical reliability.

Likely indicators

We need to distinguish between context and performance indicators. The former provide useful context to the reader so that she has some feeling for the type of institution being referred to. Obvious context indicators include:

- Measures of size (number of students, FYTEs, number of campuses)
- Age distribution
- Gender distribution
- Fields of study distribution
- Levels of study distribution
- % of full-time/part-time students
- % of students in various groups (e.g. Indigenous, disability, non-English speaking, regional distribution)

Performance indicators by contrast need to focus on outcomes. Possibilities include:

- Load pass rates
- Qualification completion rates
- Student satisfaction: overall satisfaction, teacher quality. Assessment quality, learning outcomes, reason for leaving, whether achieved main goal
- Employment outcomes: whether employed, whether employed full-time, wages, occupational status. Note that employment status before training needs to be taken into account here.
Performance indicators need to have face validity, be robust, and be able to differentiate between institutions. By face validity we mean that if institution A has a higher score than B then, everything else being equal, it will be generally accepted that institution A is performing better than B. By robustness we are referring to the amount of noise in the indicator. Thus performance indicators calculated on insufficient students will be very noisy and not robust. Finally, indicators need to be able to differentiate between institutions. There is little point in having a performance indicator which takes the more or less the same value for every institution.

The indicators we suggested above can all be derived from large scale data collections—either the students and courses collection or the student outcome survey. More radical indicators are possible if we broaden the source data. In particular, some websites allow individuals to rate various aspects of an institution, in a similar way to Amazon asking readers to rate books. This approach is quite common in the United States but probably would not be appropriate for an 'official' site.

There have also been suggestions within the AQTF that indicators of employer satisfaction should be pursued. The difficulty here is that there is no way of collecting reliable data on individual providers. While the Survey of Employer Use and Views is conducted on a regular basis, it is designed to obtain use of, and views on, various aspects of the VET system and cannot be used to provider indicators of individual provider performance.

The appropriate level of analysis

TAFEs are rather large institutions and so the idea of overall performance is problematic. Some areas within a TAFE may be much stronger than other areas. This suggests that it might be better to calculate performance indicators for fields of study, or even a field of study by qualification level. The problem with this is that the number of students at a field of study level will be a lot smaller than at a whole of institution level. This means that the performance indicators will be more robust at a whole of institution level but potentially less informative. Common websites such as the ones in the United Kingdom typically allow the user to specify the field of study. Interestingly, these websites don't seem to worry too much about the statistical properties of the indicators.

Statistical approaches

There are two basic approaches to adjust for the fact that providers have different student profiles, and these differences impact on performance indicators.

First, multilevel regression can be used to separate out the provider specific aspect of an indicator from the effect of the characteristics of students. An example is given in Curtis (2010) in which the module completion rate of individual TAFES is modelled as a function of a range of student characteristics. The control variables comprise: gender, age, Indigenous status, disability status, commencement status, location, employment status, school attainment, post-school qualifications, reasons for study, study load, whether at school, main field of education, AQF level, main delivery mode and main funding source. Curtis finds that 95% of the variation in observed module pass rates can be explained by these characteristics of students, leaving 5% of the variation which is attributed to institutional factors. Putting it another way, the observed modules pass rates are a very poor indicator of institutional performance and it would be inappropriate to use the raw scores as an indicator of performance. When Institutions are ranked by the modelled module pass-rate (i.e. abstracting from student characteristics), we find that there are bands at the top and bottom in which the providers are statistically different, but the performance of many in the middle is not statistically different from each other. Figure 1 shows
the results with the diamond representing the modelled performance relative to the overall average (the bars show statistical confidence intervals).

**Figure 1**  Profile-adjusted competency completion rates for TAFE providers

![Graph showing profile-adjusted competency completion rates for TAFE providers](image)

This graphical presentation shows why relative performance should not be converted into ranks, for the simple reason that the middle performers are very close to each other. For example, TAFEs ranked between 10 and 20 are not statistically different from each other.

The second approach is to identify peers for each provider.

There are various ways of doing this. The first is to take a naive approach and group them together on variables that have some face validity. For example, we could classify providers by size and whether they are city based or not. The obvious difficulty with this is that two providers may be similar in size but have very different orientations and student bases.

The second way is to undertake a multivariate cluster analysis which groups providers together on the basis of certain characteristics. Essentially the technique works out the distance between providers on the basis of differences between the characteristics. The advantage of this method over the first method is that it is less arbitrary and can take into account a considerable number of variables that are relevant. However, there are two disadvantages. The first is that we need to specify the relative importance of variables; otherwise every variable is treated as important as every other variable. The second disadvantage is inevitably the groupings produced by the cluster analysis will not accord with the 'common sense' priors of some stakeholders, who will then attack the validity of the technique.

The third possibility is to define peers on the basis of predicted values for a range of performance indicators, where the predictions take into account the student characteristics of each provider. The approach is to run regressions predicting the performance indicators in question. The peers of a provider are the half dozen providers whose predicted performance is
close to that of the provider in question. The idea is that the characteristics of these providers are very similar in the way that matters for the performance indicators.

The third possibility is attractive but again suffers from the difficulty that peers will not necessarily accord with ‘common sense’.

Irrespective of which method is used, it is still important to take into account the statistical variation of the performance indicator. Once we have defined a set of peers we still need to make a judgment about whether, for example, a module completion rate of 87% is significantly different from one of 85%. This will depend on how many observations are used to generate the indicator for each provider.

Final comments

Indicators in the VET sector are inevitable. There is a demand for them by consumers and by regulators. Providers should also be seeking them for benchmarking purposes. The one thing that is certain is that they will be controversial - providers who rank highly will be happy to use them but those who rank more lowly will denigrate them. To ensure that indicators are defensible it is critical that the serious statistical issues that are associated with them are addressed.

References