10th European Quality Assurance Forum

19-21 November 2015

Quality Assurance Agency and UCL Institute of Education

London, UK

Taking stock and looking forward

Paper presented during EQAF 2015

Author(s)

Name: Alex Griffiths

Position: PhD student

Organisation: King's College, London

Country: UK

Short bio:

Alex Griffiths is a doctoral student at King's College London, University of London and is the holder of an ESRC collaborative scholarship, working with the QAA to produce an empirical analysis of risk-based approaches to quality assurance in higher education. He started the project in 2013 having spent five years developing risk assessment tools and performance measures for the Care Quality Commission, England's regulator for the quality of health and social care. His PhD research is supervised by Professor Baroness Alison Wolf, and uses mathematical modelling to consider the accuracy of a range of predictive risk indicators in relation to QAA review outcomes.

Name: Dr. Elizabeth Halford

Position: Head of Research and Intelligence

Organisation: Quality Assurance Agency for Higher Education (QAA)

Country: UK

Short bio:

Elizabeth Halford is Head of Research and Intelligence at QAA. She joined QAA in 2012, after working in UK universities and colleges as an academic and senior manager. She has led programmes of teacher education and staff development, and was previously Head of Validation and Review at the University of West London. In her current role, she leads QAA research activities in applied research in the field of higher education, together with policy and intelligence functions. She holds a professional doctorate in Education from UCL IOE and her research experience includes funded projects for the ESRC and HEFC on dual-sector institutions, academic identity and widening participation.

Proposal

Title: Zen and the art of risk assessment: what are the implications of a system of risk- based quality assurance for higher education in England?

Abstract:

This paper considers how centrally-available and comprehensive quantitative data can be used as an indication of risk in a risk-based system of quality assurance, as currently implemented in England. This consideration is set within the policy context of expanding higher education and the introduction of a new system of funding undergraduate education through student loans for tuition fees in 2012.

Utilising machine learning techniques this paper demonstrates that the best model utilises three indicators relating to applications, staffing and finance. The paper concludes that the ability of data to predict the outcome of QAA reviews, and hence help prioritise them, is extremely limited.

Text of paper:

Introduction

Higher education has expanded significantly in the past decade, with systems in both developed and developing economies moving from elite to mass provision (Parry, 2003). The demand for higher education frequently outstrips the ability of the state sector to provide, resulting in an expanded private sector and mixed economy provision.

An expanding system of higher education raises questions about how it will be funded and regulated, to ensure that standards and the learning experience are maintained across a diverse and differentiated range of provision. The need to regulate in complex systems raises questions about regulatory coherence and how the various objectives of regulation can be best met in a cost effective way. One approach is the implementation of risk-based quality assurance and this paper discusses how it can be applied to the system of higher education currently in place in England, together with the potential effectiveness of different models and indicators of risk.

The English context

Government policy in England, since 2012 (*Students at the Heart of the System,* BIS, 2011), has created a market in which public subsidy for higher education is channelled through a system of loans and grants to students in institutions, charging annual tuition fees of up to £9,000 for undergraduate degrees. Student finance is administered by the Student Loans Company (SLC) and publicly funded higher education providers are subject to individual Access agreements with the Office for Fair Access (OFFA). Students in private providers are able to access loans for annual tuition fees of up to £6,000 for courses designated as eligible for support. In terms of quality assurance, this policy has been accompanied by the desire to create a level playing field for all providers and to introduce a system of risk-based quality assurance (HEFCE, 2012 and 2015), which is both proportionate and cost effective. In parallel, there has also been a recent shift in policy language, moving from references in the 2011 white paper to 'lighter touch regulation' and removing 'barriers to entry', to an emphasis on tightening standards and more robust quality assurance processes.

Risk management: indicators and modelling

Recently, there has been a significant increase in the availability of indicators concerning the management and performance of higher education institutions. This has been accompanied by calls for its use to target quality assurance activity in an intelligence-led, risk-based approach. In this research, we utilised recently-developed machine learning techniques to assess the ability of data to forecast the findings of quality assurance reviews, and hence drive a more targeted and efficient assurance system.

Background to the research

The 2011 White Paper, *Students at the Heart of the System*, called for a risk-based approach to quality assurance and for the exploration of options in which the frequency, and perhaps need, for reviews would depend upon a basket of data. HEFCE's recent consultation on future approaches to quality assessment has further raised the spectre of quality oversight driven by the monitoring of data (HEFCE, 2015a).

In 2011/12 there were 164 universities in the UK, each covered by hundreds of performance indicators, responsible for 2,500,000 HE students. This represented approximately 90% of all UK HE students (HESA, 2012; HEFCE, 2015b). Whilst the likelihood of a university receiving an 'unsatisfactory' review is low when compared with other provider types (alternative providers and further education colleges), the impact of that failure, either in terms of the number of students affected or the harm done to the reputation of UK higher education, is far greater. The findings of this study should be viewed in the context of the limited number of university reviews resulting in 'unsatisfactory' judgements.

Methodology

An efficient, risk-based approach to resource allocation, must use centrally-available and comprehensive information. To determine which indicators, if any, could have predicted the outcome of past QAA university reviews all available indicators with a feasible link to quality or quality assurance processes was sourced and assessed for its suitability, the latest data available prior to each QAA review was paired with the review outcome, and then an *elastic net* machine learning approach was used to select indicators of interest and fit the most appropriate model.

The dependent variable, the outcome that this study aimed to predict, was the overall outcome of QAA reviews. All electronically-available, complete QAA reviews were extracted from the QAA's databases and working in collaboration with the QAA, those reviews performed using a methodology comparable to the current Higher Education

Review (HER) were identified and retained with mapping to the current terminology. The final university-specific data set comprised 184 reviews concerning 139 distinct providers and dating from October 2007 to May 2014.

The number of occurrences of universities not meeting expectations is both absolutely and relatively low compared to other provider types. Indeed, no university has failed to meet expectations of the 'enhancement' question. Such low numbers can be a cause for concern. Developing a model based on too few outcomes can result in a model which is susceptible to 'overfitting' whereby the model predicts every sample perfectly having learnt not just the general patterns in the data but the unique noise of each occurrence. To limit the effect of the low number of failures to meet expectations, the dependent variable was considered at 'review', rather than 'question' level. In line with QAA terminology a review was deemed to be 'satisfactory' if all judgements were 'Meets UK expectations' or above and 'unsatisfactory' otherwise. This resulted in final data containing 13 'unsatisfactory' reviews and 171 'satisfactory' reviews.

In order to comprehensively determine which factors predict the outcome of a QAA review, it was necessary to consider as complete a data set as possible. An initial review of the HE data landscape was undertaken and complemented by discussions with HESA, QAA and the QAA's external Research Advisory Group. Deliberately not included, were indicators with no feasible link to quality or quality assurance or information prohibitively resource intensive to obtain. This resulted in 754 core indicators relating to:

- staffing
- student characteristics
- estates
- finance
- student satisfaction
- applications
- destination of leavers
- QAA concerns,
- previous review performance.

The data were gathered for each indicator from 2003/04 – 2012/13 (where applicable) and a number of change over time indicators constructed, resulting in a total of 3,639 indicators. Each review was then matched with the most up-to-date data available prior to the review for each indicator. Two sets of indicators were then considered: one where any indicators which contained missing values when paired with a review were removed, and a second where missing values were imputed where appropriate.

Utilising the *elastic net* approach which combines *ridge* and *lasso* regression to perform both model stabilisation and variable selection, we were able to determine the model which, with perfect hindsight, would have best predicted the outcome of the QAA reviews without simply over-fitting the limited data set.

Results

The best model was obtained using imputed data and contained three indicators:

- APL006_Ca1 *the one-year change in the proportion of successful applicants whose age is known who are aged 25 & above.*

- KFI020_Abs *the percentage ratio of contribution from research grants & contracts to research grants & contracts income* calculated as:

$$\frac{\text{Research grants \& contracts Income} - \text{Total research grants \& contracts expenditure}}{\text{Research grants \& contracts Income}}$$

- STA062_Ca1 *the one-year change in the proportion of full-time equivalent (FTE) staff who are principally financed by the institution.*

Specifically, the model determines the probability of a university receiving an 'unsatisfactory' review as:

$$P(Unsatisfactory) = \frac{e^{(-2.61\ +\ (5\times APL006\_\mathrm{Ca1})+(-0.000088\times KFI020\_\mathrm{Abs})+(11.16\times \mathrm{STA062\_Ca1})}}{1+e^{(-2.61\ +\ (5\times APL006\_\mathrm{Ca1})+(-0.000088\times KFI020\_\mathrm{Abs})+(11.16\times \mathrm{STA062\_Ca1})}}$$

As the coefficient (the number by which the indicator is multiplied by in the above equation) is positive for the applications and staffing indicators , positive values of each indicator lead to increases in the predicted probability of an 'unsatisfactory' review whilst negative values lead to decreases. Therefore, ceterus paribus, an increase in the proportion of successful applicants (whose age is known) who are aged 25 & above will lead to an increase in the predicted likelihood of being judged 'unsatisfactory', as will an increase in the proportion of staff who are principally financed by the institution. As the coefficient is negative for the finance indicator the opposite is true: negative values, indicating a university spending more on research than the income it has received for that purpose in a given year, will lead to increases in the predicted probability of an 'unsatisfactory' review whilst positive values will lead to decreases .

Three hyopthetical universities whose performance, and the resulting predicted probability of being judged 'unsatisfactory' are considered in table 1 below:

- University A has had an increase in the proportion of succesful applicats aged 25 and over from 10% in the previous year to 20% this year, has spent double its allocated research funds, and has had an increase in the proportion of full-time equivalent staff who are principally financed by the institution from 5% last year to 10% this year.

- University B has remained exactly the same and has spent its full research budget on research; no more and no less.

- University C has had a decrease in the proportion of succesful applicats aged 25 and over from 10% in the previous year to 5% this year, has spent half its allocated research funds, and has had a decrease in the proportion of full-time equivalent staff who are principally financed by the institution from 5% last year to 2.5% this year.

Univ.	APL006_Ca1	KFI020_Abs	STA062_Ca1	Probability of being judged 'unsatisfactory'
A	0.10	-1	0.05	$\frac{e^{(-2.61+(5\times0.1)+(-0.000088\times -1)+(11.16\times 0.05))}}{1+e^{(-2.61+(5\times0.1)+(-0.000088\times -1)+(11.16\times 0.05))}} = 0.1748$

B	0	0	0	$\frac{e^{(-2.61+(5\times0)+(-0.000088\times0)+(11.16\times0))}}{1+e^{(-2.61+(5\times0)+(-0.000088\times0)+(11.16\times0))}} = 0.0685$
C	-0.05	0.5	-0.025	$\frac{e^{(-2.61+(5\times-0.05)+(-0.000088\times0.5)+(11.16\times-0.025))}}{1+e^{(-2.61+(5\times-0.05)+(-0.000088\times0.5)+(11.16\times-0.025))}} = 0.0415$

Table 1: Hypothetical values of APL006_Ca1, KFI020_Abs and STA062_Ca1 indicators and the resulting predicted likelihood of being judged 'unsatisfactory'.

In the above example University A has a 17.48% predicted likelihood, or roughly 1 in 5 chance, of being judged 'unsatisfactory' whereas University C has just a 4.15% predicted likelihood, or roughly 1 in 24 chance, of being judged 'unsatisfactory'.

Conclusions

Using only naturally-complete data no model performed better than simply assuming all universities had an equal chance of being 'unsatisfactory'. Once the data was imputed we were able to obtain a single model which utilises three specific indicators regarding successful applications, finance and institutional staffing characteristics, to predict the likelihood of being judged as 'unsatisfactory'. How well this model performs depends on one's viewpoint and risk appetite. Within the first six reviews the model would have prioritised four 'unsatisfactory' judgements; however, performance declined markedly after this point. Whilst the model would have, with perfect hindsight, required for just over half the reviews actually conducted, to identify all but one of the 'unsatisfactory' institutions, it would have still had a required 95 reviews of 'satisfactory' institutions to have been carried out. To have prioritised all 'unsatisfactory' reviews 174 reviews, including 161 reviews of 'satisfactory' institutions would have been required. Furthermore, the model's application to more recent data, not yet followed-up by reviews, suggests that some institutions which are highly ranked in global league tables, should be prioritised as 'at risk', if this type of modelling is used. Whether this means that there are high-profile institutions facing quality assurance challenges in a complex and fluid external environment, or, despite perfect hindsight and a wealth of data, we are unable to predict the outcome of regulatory reviews, is a matter for discussion.

Discussion

Considering over 3,600 indicators of university performance the best model we are able to obtain for predicting an 'unsatisfactory' review is:

$$P(Unsatisfactory) = \frac{e^{(-2.61\,+\,(5\times APL006\_\mathrm{Ca1})+(-0.000088\times KFI020\_\mathrm{Abs})+(11.16\times \mathrm{STA062\_Ca1})}}{1+e^{(-2.61\,+\,(5\times APL006\_\mathrm{Ca1})+(-0.000088\times KFI020\_\mathrm{Abs})+(11.16\times \mathrm{STA062\_Ca1})}}$$

As detailed above, there are significant concerns about the effectiveness of this model. Its application to the 2012/13 data suggests that, whilst it is doing a reasonable job of describing the historic data used to develop it, this is not translating into effective predictions based on more recent data. Even if its predictions based on the 2012/13 data are accurate, the model will still face fierce criticism for several reasons:

1. A high number of established universities, none of whom have ever received any judgement other 'Meets UK expectations' or 'Commended', are predicted as being amongst the most likely to be judged 'unsatisfactory'.

2. With the 92.5% error rate that would have occurred had the QAA utilised the model perfectly, any university would rightly assume it was highly unlikely they had been correctly prioritised when told they were going to be subject to a review.

3. It is not entirely apparent looking at the indicators contained in the model why it should be able to foretell quality assurance failures. An increase in the value of the staffing indicator could feasibly (although tenuously) serve as a proxy for flagging institutions who are changing their workforce which creates disruption and introduces a lack of continuity. There are many possible challenges to this explanation however. The indicator is based on relative numbers and does not account for base values; for example, if two universities both had an equal number of students and one doubled its staff, principally financed by the institution, from 500 to 1,000 and the other doubled its staff from 5,000 to 10,000 they would both be predicted as equally likely to be 'unsatisfactory'. This is despite the fact that the latter university would have ten times as many experienced staff familiar with the existing quality assurance processes. An alternative explanation for an increase could be that a university is maintaining its staffing numbers but that these staff are not as able to attract research funding as they were in the past and so are now reliant on the institution. If so, this would show a similar outcome to KFI020_Abs; however, it is difficult to see how either relates to one or more of the four areas covered by a QAA review: academic standards, the quality of teaching and learning, enhancement or the provision of information. Likewise, it is difficult to fathom how an increase in the proportion of undergraduate students whose age is known and who are aged 25 or over could feasibly serve as an indicator of performance in these four key areas.

The best model is therefore not only questionable when applied to the 2012/13 data, but it is not intuitive either; there is no obvious reason why a model comprised of the three indicators would have the ability to forecast the outcome of QAA reviews. It therefore appears we are unable to successfully predict the outcome of QAA reviews for universities. Despite a comprehensive analysis of thousands of indicators including past performance and student satisfaction measures and covering both absolute performance and changes over time, the best model is still highly questionable and providers are unlikely to be convinced by it.

The result gives rise to one obvious question: why, with so much data available, are we unable to predict the outcome of QAA reviews for universities? There are six logical possibilities as shown in figure 1 below:

EUROPEAN QUALITY ASSURANCE FORUM

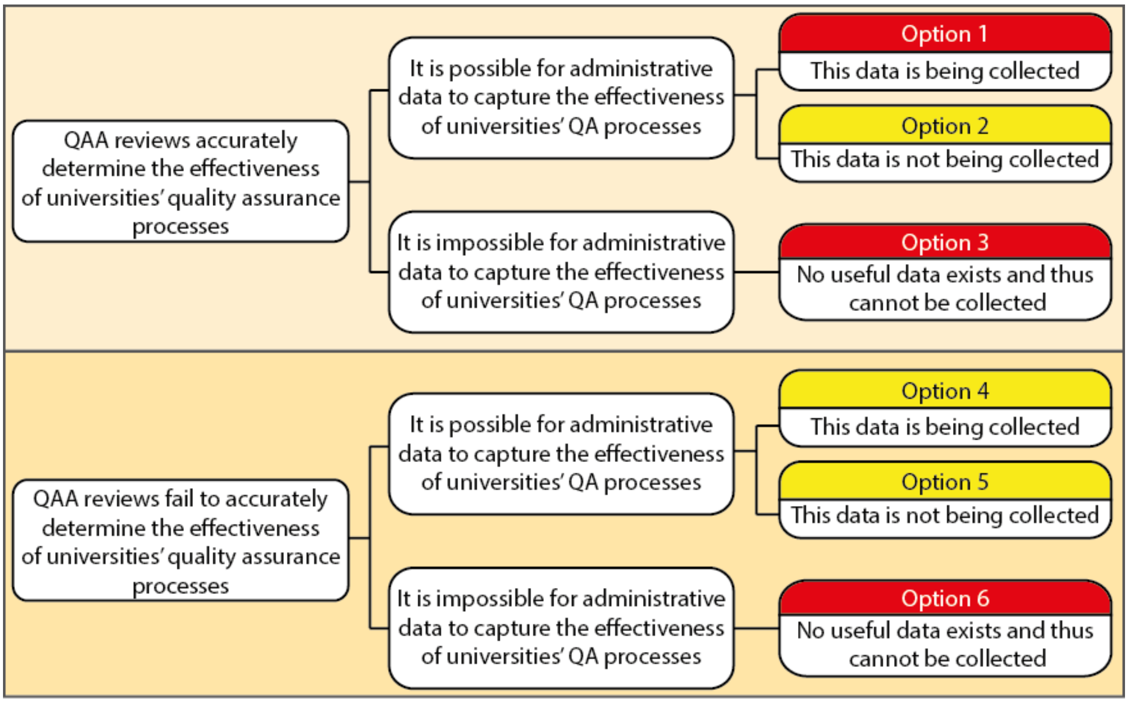

Figure 1: Possible reasons for our inability to predict the outcome of QAA university reviews.

This study has shown that option 1 - QAA reviews are an accurate assessment of each university's QA processes, data can capture that process, and it is being collected - is not true. We have considered all those indicators which could feasibly be used in a risk-based approach by the QAA and have demonstrated the best possible model is not good enough. If option 2 represents reality - that QAA reviews are effective, data can capture whether or not QA processes are effective, but these data are not being collected - means a risk-based approach could be successfully operated in future with additional data gathering. This, however, seems unlikely. This study has considered over 3,600 indicators covering a broad range of topics including past performance, student satisfaction, staff and student characteristics, applications and finance. It is not obvious what additional data could be collected.

The remaining two options, options 4 and 5, relate to the scenarios where QAA reviews are ineffective, they are not accurately measuring the effectiveness of QA processes, but appropriate data could be incorporated into a burden-reducing, risk-based approach. If the data is being collected (option 4) there is nothing that can be done with it as we are trying to predict something which we do not know, we have no way of knowing which universities are having quality assurance issues other than when they are significant enough to garner media attention and even then it would be difficult to judge the threshold for what actually constitutes a quality assurance failure. If, as in option 5, the data exists to capture the effectiveness of QA processes but the QAA's reviews are ineffective and the data is not being collected we are again unable to implement an effective risk-based approach.

Implications for policy and practice

When considering the implications for policy and practice in higher education of a system of risk-based quality assurance, it is necessary to ask what constitutes quality in higher education and what are the objectives of quality assurance in particular systems? These may include:

- Ensuring accountability for public investment
- Assuring academic standards
- Protecting the student experience
- Promoting enhancement

The title of this paper alludes to *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values* (Pirsig,1974), in which the meaning and concept of quality, a term he deems to be undefinable, is explored. It is suggested that to truly experience quality one must both embrace and apply it as best fits the requirements of the situation. This approach would support the importance of a regulatory system designing methods of review, audit or accreditation that are fit for purpose for a wide range of different types of higher education provider, with different missions, purposes and traditions, as a necessary and contingent feature of a diverse system of higher education. This can be addressed by the concept of threshold standards and risk-based review which are more proportionate and cost effective, reflecting the levels of risk at different institutions, based on both retrospective and predictive indicators.

References:

Department for Business, Innovation and Skills (2011). *Higher Education: Students at the Heart of the System*. Cmnd 8122. DBIS: London.

Higher Education Funding Council for England (2012). *Invitation to the Quality Assurance Agency to Implement a more risk-based approach to the quality assurance of higher education in England*, letter to Anthony McClaren, Chief Executive of QAA, from Alan Langlands, Chief Executive of HEFCE, 6 Nov.

Higher Education Funding Council for England (2014). *Review of Quality Assessment.* Available at: http://www.hefce.ac.uk/reg/review/

Higher Education Statistics Agency (2013). *General Student Numbers* [online]. Available at: http://www.hesa.ac.uk/content/view/1897/239/

Parry, G. (2003) *Mass Higher Education and the English: Wherein the Colleges?* Higher Education Quarterly, 0951-5224 Volume 57, No. 4 pp 308-337

Pirsig, R.M. (1974) *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values,* William Morrow: USA

Quality Assurance Agency (2011). *Institutional Review of Higher Education Institutions in England and Northern Ireland: A Handbook for Higher Education Providers*, QAA: Gloucester.

HEFCE. (2015a). Future Approaches to Quality Assessment in England, Wales, and Northern Ireland - Consultation [Online]. Available: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2015/201511/2015_11_.pdf [Accessed 20/07/2015].

HEFCE. (2015b). Personal Correspondence with the author 24/03/2015.