

Written Evidence from Alex Griffiths

Contributor

1. Alex Griffiths is a researcher at King's College London working with King's Centre for Risk Management (Dr Henry Rothstein) and its International Centre for University Policy Research (Professor Alison Wolf) to conduct an empirical analysis of risk-based approaches to quality assurance in higher education. This research is funded by the Economic and Social Research Council (ESRC) and is carried out in collaboration with the Quality Assurance Agency (QAA). The views expressed in this evidence are his personal views arising from this research and do not necessarily reflect those of King's College London or QAA.

Summary

2. The Quality Assessment Review Consultation published in June 2015 suggested that the new quality assessment system should be "proportionate and risk-based and should, where possible, minimise the burden and cost on providers by making use of existing data and information."¹
3. This is an attractive goal. High-quality providers would be freed from the burden of unnecessary regulation whilst low-quality providers would be given the additional regulatory attention they need. There is, however, no empirical research on what data could be used to determine which providers are most likely to be 'unsatisfactory'².
4. To address that gap, we at King's performed a comprehensive analysis of hundreds of QAA reviews and thousands of contemporaneous indicators to determine whether it is possible to build a robust model that could use existing information to predict the likely outcome of QAA reviews. Without such a model to prioritise higher education providers for review, a risk-based approach cannot work.
5. The analysis could not identify any meaningful relationship between the vast array of higher education data and the subsequent outcome of QAA reviews. It was, therefore, concluded that no satisfactory model could be built using available data and information. Any available model would instead result in high-quality providers being unnecessarily burdened and stigmatised and low-quality providers going undetected.

Background to the Research

6. The 2011 higher education white paper *Students at the Heart of the System* called for the Quality Assurance Agency (QAA) to adopt "a genuinely risk-based approach", focusing effort where it will have most impact, and to "explore options in which the frequency – and perhaps need – for a full, scheduled institutional review will depend on an objective assessment of a basket of data, monitored continually but at arm's length."³

¹ Quality Assessment Review Steering Group. (2015). *The Future of Quality Assessment in Higher Education*, Paragraph 10.

² See for example: Raban, C. (2011). Talking About Quality: Risk and Regulation. *The Quality Assurance Agency for Higher Education, Gloucester*; King, R. (2011). *The Risks of Risk-Based Regulation: The Regulatory Challenges of the Higher Education White Paper for England*, HEPI.

³ BIS (2011). *Students at the Heart of the System*. Cmnd 8122. London: BIS.

7. HEFCE subsequently consulted on a proposal to introduce an annual assessment of nationally collected data that could prompt QAA reviews⁴. The proposal received strong negative feedback. 71% of responses expressed concerns about the scope, validity, availability and reliability of data.⁵ The proposals were, consequently, not implemented.⁶
8. 'Risk-based' approaches continued to be discussed in this and other contexts, and their attractions were and are clear. King's College London obtained funding and QAA collaboration for research into the feasibility and likely accuracy of moving to such an approach, and work commenced in autumn 2013.
9. In January 2015 the Quality Assessment Review Steering Group led by HEFCE revived the idea of quality assurance reviews being prompted by national data sets. In a 'discussion document', the Steering Group sought views on, amongst other things, the principle that future quality assessment should be "risk-based but responds swiftly where problems are identified and provides for proportionate, escalatory action where necessary".⁷
10. The June 2015 consultation that followed stated there was broad consensus that any future "quality assessment system should be proportionate and risk-based and should, where possible, minimise the burden and cost on providers by making use of existing data and information."⁸
11. To date, however, there have been no published analyses of which indicators, if any, are effective predictors of peer-review quality assessments in higher education. For example, it is not known whether the indicators suggested in the June 2015 consultation, including the National Student Survey (NSS) and Destinations of Leavers from Higher Education (DLHE) survey, are of any use in helping target poor quality provision as part of a risk-based approach. Our research, for which the QAA has provided us with full data from 1999 onwards, is the only analysis making use of all these data: and we are not aware of any other in-depth empirical studies in this field. Our own analyses were completed in summer 2015 and are therefore as up-to-date as is feasible.

Research Design

12. The research uses the fact that we have had full reviews for nearly all institutions to investigate whether those providers who were judged unsatisfactory after a full review could have been identified in advance using available data. If so, then, in principle, a data-driven, risk-based approach to quality assurance could have been used in the past: and the research findings could play a key role in informing future risk-based approaches and developing workable metrics. If it proved impossible to identify high-risk providers, the research would suggest that any future risk-based approach is unlikely to succeed.

⁴ HEFCE. (2012). *A Risk-Based Approach to Quality Assurance: Outcomes of Consultation and Next Steps*

⁵ HEFCE. (2012). *A Risk-Based Approach to Quality Assurance: Outcomes of Consultation and Next Steps Annex A: Detailed Analysis of Responses*.

⁶ Langlands, A. 6 Nov 2012. *RE: Invitation to the Quality Assurance Agency to Implement a More Risk-Based Approach to the Quality Assurance of Higher Education in England*. Letter to McLaren, A.

⁷ Quality Assessment Review Steering Group. (2015). *The Future of Quality Assessment in Higher Education*.

⁸ HEFCE. (2015). *Future Approaches to Quality Assessment in England, Wales, and Northern Ireland - Consultation*

13. The research looked at three higher education sectors – universities, higher education in further education colleges (FECs), and alternative providers – separately to make full use of the available data.
14. All electronically-available QAA reviews that were comparable with the current QAA methodology were extracted from the QAA's databases in late November 2014. The final data set comprised 858 reviews concerning 699 distinct providers undertaken between October 2007 and November 2014.
15. For each sector all available performance data that could feasibly form part of a risk-based approach, i.e. is available for a large proportion of higher education providers and is not prohibitively expensive to gather, was sourced and assessed for its suitability. For universities these data included, NSS, DLHE, staffing, students, finance, HESA performance indicators, QAA Concerns, past review performance, research, and applications. For FECs, where less data is available, this included financial information, staffing and student characteristics, Ofsted ratings, QAA Concerns and past review performance. For alternative providers, where even less data is available, some 600 sets of financial accounts were purchased from Companies House and considered alongside QAA concerns and past review performance measures.
16. For each of the hundreds of indicators change-over-time variants were calculated where possible so that each provider's 'direction of travel' could be considered in addition to its current performance. The latest data available prior to each QAA review was then paired with the review outcome.
17. Consideration was given to whether a proportion of the reviews from each sector could be withheld from the model development phase to test the predictions of the respective models once developed. For universities, the limited number of reviews, and more importantly the limited number of 'unsatisfactory reviews – just 13 out of 184, meant that all reviews were needed to develop a robust model. Instead, to sense-check the model the predictions made using 2012/13 data were compared to the Guardian University Rankings for that time period⁹. For FECs the limited number of reviews also prevented us from withholding data to test the model; however, ongoing review activity by the QAA meant a new tranche of review findings was available on which to test the model. For alternative providers there was a sufficient number of reviews available to withhold a representative sample of 20% of reviews for testing any model.
18. A regression analysis was run for each individual indicator to gain an early insight into which may be related to the outcome of QAA reviews and to what degree. These results were informative but should be treated with caution. If you compare enough unrelated data sets you will always find strong correlations between individual variables: but such correlations do not in themselves indicate that there is any true causal or generalizable relationship them. For example, between 2000 and 2009 in the USA per capita consumption of mozzarella cheese correlated near perfectly with the number of civil engineering doctorates awarded¹⁰.
19. Iterative machine-learning techniques were then used to fit the most appropriate model.

⁹ The Guardian. (2013). University Guide 2014: University League Table [Online]. Available: <http://www.theguardian.com/education/table/2013/jun/03/university-league-table-2014> [Accessed 24/10/2015].

¹⁰ Vigen, T. (2014). Spurious Correlations [Online]. Available: <http://www.tylervigen.com/> [Accessed 13/09/2014].

20. Success could be judged in two ways. If the preference is for no 'unsatisfactory' providers to escape being prioritised by the model, then all 'unsatisfactory' providers must be identified as high-risk without also identifying, and therefore burdening with additional reviews, too many 'satisfactory' providers; otherwise there is little benefit to a risk-based approach. If we accept some undetected 'unsatisfactory' provision is inevitable and acceptable, then the model should aim to identify a group of 'high-risk' providers that turn out to be 'unsatisfactory' more often than those providers in the 'low-risk' group. The preferred measure of success, and the acceptable rate of false positives ('satisfactory' providers identified as high risk) and false negatives ('unsatisfactory' providers identified as low risk), is a matter of policy. The results for both measures are detailed below.

Findings

21. While the results differ for universities, FECs, and alternative providers, overall the picture is very clear: those providers who were judged 'unsatisfactory' after a full review could not have been reliably identified in advance using available data. This suggests that any future data-driven, risk-based approach is unlikely to succeed.

Universities

22. Whilst a variety of the 1,690 indicators examined correlated with the outcome of QAA reviews there was no greater number of such indicators than we would expect to see by chance alone. Many of those indicators that did correlate with past review outcomes appeared to be useful predictors of 'satisfactory' universities, in the sense that universities with a very high score on an indicator would nearly always be 'satisfactory'. Very low scores on these same indicators did not, however, mean that a university had received an 'unsatisfactory' judgement.
23. Multiple iterations provided us with a model which did the best job possible of identifying universities that had been 'unsatisfactory' without using very large numbers of indicators. The indicators were chosen for their statistical power, not because of any apparent causal link. The final model contained three indicators: one relating to the change in the proportion of successful applications from mature students (i.e. the greater the increase in the proportion of mature students accepted, the higher the likelihood of being judged 'unsatisfactory'), one relating to research spending (i.e. the greater the proportion of research budget overspend, the higher the likelihood of being judged 'unsatisfactory'), and one relating to the financing of staff members (i.e. the greater the increase in the proportion of staff funded primarily by the university, the higher the likelihood of being judged 'unsatisfactory').
24. This model was not able to identify all 'unsatisfactory' universities. If reviews had been conducted in turn, from the most 'at risk' on this model on down, it would have required 174 out of a total of 184 possible reviews to have been conducted to successfully identify all the 'unsatisfactory' universities.
25. If reviews had been conducted in turn, from the most 'at risk' on this model on down, 11 out of the 13 'unsatisfactory' universities would have been identified by the time half of the 184 reviews had been conducted. This would have meant that, if allowing two 'unsatisfactory' universities going undetected were considered acceptable, 92 reviews could have been forgone

by the QAA. 81 out of the 92 reviews still conducted would have resulted in a 'satisfactory' judgement.

26. However, when the model was applied to new data from the 2012/13 academic year it produced some very questionable results including prioritising three of the UK's top 10 universities amongst the most likely to be judged 'unsatisfactory'. This may be because there are a number of highly-ranked trends in higher education have unknown quality assurance issues or, more likely, because the apparent associations between the data and the review outcomes in the past were merely coincidental.

FECs

27. As with universities there were a number of indicators for FECs that had a strong correlation with the QAA review findings, but no more than we would expect to see by chance alone.
28. The final model contained two finance indicators and one indicator relating to the outcome of the FEC's previous comparable QAA review. Counter-intuitively, those FECs who had previously had a positive QAA review comparable to the current methodology adopted by the QAA were more likely to be judged 'unsatisfactory' than those who had not. This is likely to be a result of the cyclical nature of reviews which means that those FECs who have had a previously comparable review have also been reviewed more recently under the QAA's tougher *HER* approach and at a time when FEC budgets are under considerable pressure.^{11 12}
29. Use of this model, with inspections being carried out in 'descending order' of risk, would have required 102 reviews out of a total of 131 reviews to have been conducted to successfully identify all the 'unsatisfactory' FECs.
30. It would have placed 18 out of the 20 'unsatisfactory' FECs in the riskiest half of FECs. If allowing two 'unsatisfactory' FECs going undetected were considered acceptable, 66 reviews could have been forgone by the QAA. 47 out of the 65 reviews still conducted would have resulted in a 'satisfactory' judgement.
31. However, when applied to new QAA review findings the model performed even worse than tossing a coin. The QAA would have been better off doing the exact opposite of what the model suggested and prioritising the *low* risk FECs. Any tenuous relationships which existed between the data and the QAA review outcomes in the past no longer hold. This may be because trends in higher education have shifted or, more likely, because the apparent associations between the data and the review outcomes were merely coincidental.

Alternative Providers

32. Three indicators strongly correlated with the outcome of QAA reviews: the provider's age at the time of their review, the amount of 'Cash at bank and in hand' and their Total Net Assets. Again, there was no greater number of highly correlated indicators than we would expect to see by chance alone and these predictors were more useful at predicting which providers were highly likely to be 'satisfactory' rather than 'unsatisfactory'.

¹¹ Wolf, A. (2015). *Heading for the Precipice: Can Further and Higher Education Funding Policies Be Sustained?* [Online]. The Policy Institute at King's.

¹² QAA. (2014). *Higher Education Review: First Year Findings 2013-14*

33. The final model contained indicators relating to the size of the provider, past review performance and their financial position. Larger, established providers in a strong financial position who had previously undergone a QAA review were least likely to be 'unsatisfactory' whereas smaller, younger providers who had not previously undergone a QAA review were most likely to be 'unsatisfactory'.
34. Use of this model, with inspections being carried out in 'descending order' of risk, would have required 80% of all reviews to have been conducted to successfully identify all the 'unsatisfactory' alternative providers.
35. It would have placed 28 out of the 33 'unsatisfactory' alternative providers in the riskiest half of alternative providers. If allowing five 'unsatisfactory' alternative providers going undetected were considered acceptable, 118 reviews could have been forgone by the QAA. 89 out of the 117 reviews still conducted would have resulted in a 'satisfactory' judgement.
36. When applied to QAA review findings withheld for testing, the model performed marginally better than chance although this improved performance was not statistically-significant.

Conclusion

37. The rationale for a risk-based approach is attractive. However, it cannot be achieved in practice. Our comprehensive research has considered hundreds of QAA reviews and sourced and analysed thousands of contemporaneous indicators. No effective model could be developed for identifying 'unsatisfactory' providers and prioritising them for review.
38. If it is considered unacceptable to allow any 'unsatisfactory' provision to go undetected then, for all intents and purposes, every provider will need to be reviewed and a cyclical approach to reviews would be far simpler and cheaper to adopt.
39. If it is considered acceptable to allow some 'unsatisfactory' provision to go undetected then the approach may be feasible. The best models can separate a 'high-risk group' with a greater proportion of 'unsatisfactory' providers than the 'low-risk group' with a smaller proportion of such providers. However, while the best models can achieve this retrospectively, when applied to new data, they perform very poorly.
40. Breakdown in the models, when used for new data, means that any relationships that existed between the data and the QAA review outcomes in the past no longer hold. This may be because trends in higher education have shifted or, more likely, any previous associations between the data and the review outcomes was purely coincidental.
41. This result does not mean that either the data or the review outcomes are inaccurate. Rather, they are measuring different things. Using one to try and predict the other cannot work and we therefore conclude that, in the case of quality assurance in higher education, a risk-based approach using available data is not feasible.