



13th European Quality Assurance Forum

Broadening the scope of QA

Hosted by WU (Vienna University of Economics and Business) and AQ Austria
15-17 November 2018

Call for contributions: paper submission form

Deadline 24 July 2018

Please note that all fields are obligatory. For a detailed description of the submission requirements and Frequently Asked Questions please consult the Call for Contributions.

ISSN: 1375-3797

Author(s)

Name: Helen Cullis (corresponding author and presenter at the Forum)

Position: Data Analyst

Organisation: Quality Assurance Agency for Higher Education

Country: United Kingdom

E-mail address: H.cullis@qaa.ac.uk

Short bio (150 words max):

Helen Cullis is the Data Analyst at the Quality Assurance Agency for Higher Education (QAA) in the Evaluation and Analytics team. After graduating from the University of Oxford with a degree in human sciences Helen has worked mainly in the charity and public sectors. She has worked at QAA since early 2017 and brought with her the learnings and experience of data management and research in medical research from her time at the Medical Research Council Clinical Trials Unit at UCL and at Gloucestershire Hospitals NHS Foundation Trust.

Helen manages QAA's research portfolio including any commissioned research projects. She is a key member of the Data Information Governance Group working to improve QAA's data capability and management; collaborating with sector colleagues to achieve this. The team provides intelligent, thoughtful and relevant evaluation, analysis, policy and research, producing reports, papers and other relevant work for the benefit of the sector.

Name: Alex Griffiths (co-author)

Position: Director of Data Science

Organisation: Statica Research

Country: Netherlands

E-mail address: Alex.Griffiths@StaticaResearch.com

Short bio (150 words max):

Alex is Director of Data Science at Statica Research, a company he co-founded while working as a Researcher at the London School of Economics' *Centre for Analysis of Risk and Regulation*. Having spent eight years working for the Care Quality Commission, the regulator of health and social care in England, and completed his PhD applying machine-learning techniques to improve the prioritisation of



higher education quality assurance reviews, Alex has pioneered the use of online user feedback to monitor the quality of public services, including healthcare and higher education.

If there are several authors, please copy and fill in the fields for each author and indicate who is the corresponding author and who will be responsible for presenting the paper at the Forum.

If you are submitting a paper or workshop proposal, please do not register for the event online until the results of the selection process have been announced. Each selected paper and workshop at EQAF 2018 will benefit from one reduced fee, which will be applied through a special registration process.

During the Forum, the full text of all papers presented at the Forum as well as the associated Powerpoint presentations will be published on the Forum website. If you do not wish your paper to be published, please indicate so here. This has no consequences on the selection of the papers. Please however note that all Powerpoint presentations will be published, regardless of whether the full paper is published.

Proposal

Title: The Wisdom of Students: Monitoring Quality via Social Media and Student Reviews

Abstract (150 words max):

Many of the student feedback surveys at provider and sector level give a comprehensive insight into the student experience, but they offer only a snapshot at set points in time, often at the end of a term or year. Building on innovative research in the healthcare sector, where real-time patient feedback via social media and other online feedback sites has been shown to effectively predict the outcome of in-depth inspections, this paper explores the value of continuous monitoring of the student experience via social media and other sources. What insights can be gathered from year-round feedback, what do students actively feedback on, and what relation does this feedback have with other quality measures? Could this approach be used in higher education to broaden the scope of quality assurance and enhance quality assurance systems by encompassing a broader range of activities into those systems?

Has this paper previously been published/presented elsewhere? If yes, give details.

This paper is an updated version of a piece of research commissioned by the Quality Assurance Agency for Higher Education (QAA), which is published on the QAA website (<http://www.qaa.ac.uk/en/about-us/what-we-do/policy-and-research/research>). This paper includes data from an additional source. If invited to present at the Forum we would include a progress update on the pilot project (follow-up research) running in autumn 2018. The original version of the research was presented in the UK at QAA's Annual Conference in April 2018 and at a Quality Strategy Network Event in July 2018.

Text of paper (3000 words max):

Introduction

The English higher education sector is very diverse with different organisation types delivering higher education (universities, colleges and a wide variety of other institutions, charities and companies – all collectively known as providers):

- Higher education institutions (HEIs) – these are mainly universities
- Further education colleges (FECs) - their core business is typically level four on the European Qualifications Framework but they also act as providers of HEI level qualifications (which are typically validated by an HEI) and also offer progression to higher levels of learning at another provider
- Alternative providers - these include for (or not for) profit, charitable and private providers.

Over the last two decades government policy for higher education in England has focused on the marketisation of the sector. Current policy places students at the centre of a consumer-led system, where better-informed student choice between providers is intended to increase competition, quality and learning innovation. To support this choice and improve quality there has been an increase in the collection and availability of course, and provider-level, performance data. The results of large national data collections, such as the National Student Survey (NSS) and Destinations of Leavers of Higher



Education (DLHE) survey, are widely reported: these data inform multiple league tables that aim to rank providers and inform student choice. Regulators too have made use of this outcome data to support student choice and promote good-quality provision. The Teaching Excellence and Student Outcomes Framework (TEF), for example, uses benchmarked data on students' entry qualifications, student satisfaction, and employment, to broadly grade providers. according to student performance and student outcomes.

Regulators' use of data has not been limited to teaching excellence however. In line with broader changes to regulatory practices, external oversight bodies have aimed to reduce the burden on (the great majority of) providers through data-driven, risk-based approaches. The Office for Students (OfS) (previously the Higher Education Funding Council for England (HEFCE)) has also sought to regulate using lead indicators and periodic, but not cyclical, assurance judgements based on risk analyses of these data.

A key challenge to the success of this new regulatory environment is to ensure that data is timely, robust, low-burden and representative of student views and experiences. Several problems stand in the way of obtaining such data. First, shaped by the cyclical nature of higher education, existing data collections are annual and retrospective, providing a 'lagged' view on provision that make timely interventions more challenging. Second, data must be collected by the providers whose performance will be judged on that data. This leaves the risk of providers 'gaming the system' in ways that may impact on the accuracy of the data submissions and sector-wide comparisons. Third, it is costly, time consuming and burdensome to develop new data collections geared to the specifics of a new system. Consequently, existing data collections tend to be adapted to a purpose quite different from their original design. Fourth, most existing data collections measure what the designers of those measures deemed important, rather than what students in their growing role as consumers deem important.

One potential solution to the challenge of securing timely, robust, low-burden and insightful data may come from the student body. Francis Galton discovered the 'wisdom of crowds' phenomenon¹ due to a competition to guess the weight of a cow at a local fair. There were 800 competitors, most of whom were not experts, who submitted their guesses on numbered cards. Curious about the entrants' guesses, Galton borrowed the entry slips once the competition was over and the weight of the cow was found to be 1,198 lbs. Much to his surprise, the average of the entrants' guesses was 1,197 lbs, essentially perfect. The 'Wisdom of Crowds' phenomenon that Galton had discovered means that, under the right circumstances, groups can be remarkably insightful. This can be the case even if the majority of people within a group are not especially well informed or rational. Whilst we as individuals rarely have all the necessary facts to make an accurate assessment, and are subject to numerous biases, when our individual assessments are combined in the right way, our collective assessment is often highly accurate.

This phenomenon has already been shown to be effective in healthcare in England where the collected online feedback and social media postings of patients and their families has been proven to be a statistically-significant predictor of the outcome of subsequent in-depth quality inspections². The research presented in this paper was commissioned by the Quality Assurance Agency for Higher Education (QAA) to explore whether student reviews align with more formal quality assurance measures in the higher education sector.

Data landscape

There is a significant volume of student feedback available online. For this research, over 230,000 reviews were gathered from 165 higher education institutions, 211 further education colleges and 12 alternative providers in the UK considered in scope. These reviews came from four sources: Facebook, Google, Whatuni.com and Studentcrowd.com. Facebook data was the only source that contained a significant number of irrelevant comments that we attempted to clean from the data. The significant volume of reviews available online is a relatively new phenomenon. As Figure 1 shows, the number of student reviews made each year, using the four data sources considered as part of this study, has grown significantly since 2012. By the end of February 2018, over 16,500 further reviews had already been posted.

Count of Reviews by Year

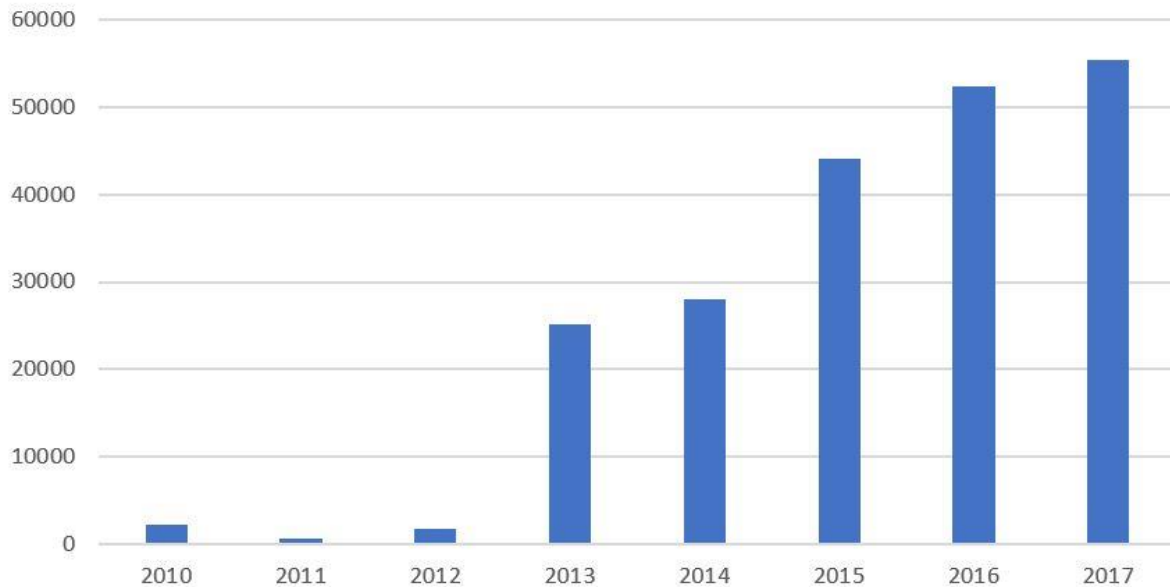


Figure 1: The number of reviews available by year from 2010 – 2017.

For each of the four data sources users are required to rate their overall experience on a scale of one (worst) to five (best) stars (Table 1). The scores can relate to the reviewer's overall experience of the provider, or their opinion of a specific course, department or set of facilities within a provider. Reassuringly for the sector, the average score over the 230,000 reviews is 4.18 stars, suggesting a high level of overall satisfaction with UK higher education.

	2014	2015	2016	2017
HEI	4.23	4.25	4.23	4.20
FEC	4.18	4.07	3.93	3.81
AP	4.05	4.33	3.97	3.82
Combined	4.23	4.24	4.21	4.16

Table 1: The average review score by provider type and year.

This high level of satisfaction has remained consistent over the past four years and across the three types of higher education provider. Caution should be taken when considering the fall in the average review score for APs given the limited number of reviews - just 392 in 2017. The average review score also varies across the year as can be seen in Figure 2.

Average Review Score by Month

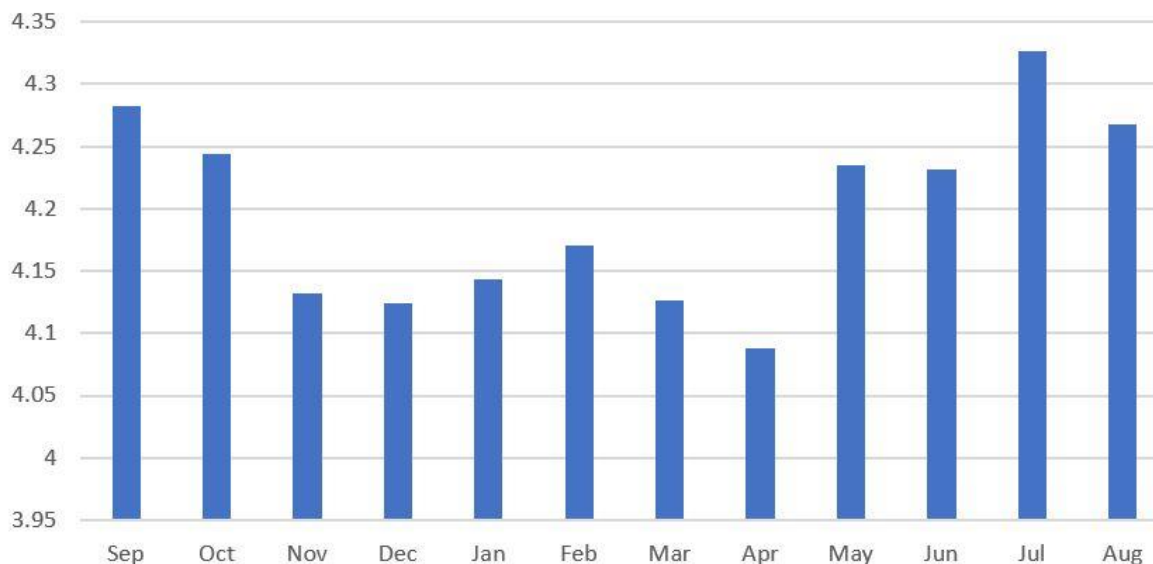


Figure 2: The average review score by month across all provider types.

Other measures of quality

In this research the average review scores from the four data sources were compared against three formal quality measures: Annual Provider Review (APR), The Teaching Excellence and Student Outcomes Framework (TEF) and The National Student Survey (NSS).

Annual Provider Reviews

APR was introduced in 2016-17, it considers existing data that is used in a 'contextualised and rounded way' with judgements on quality and standards matters reached through peer review³. Of the 376 HEIs and FECs included in this study, 330 had a HEFCE APR rating as of 20 March 2018. Providers are judged as 'meets requirements' with no further action required or 'meets requirements with an action plan', meaning an action plan was required to address areas of immediate concern.

Teaching Excellence and Student Outcomes Framework

The TEF was designed by the Department for Education to assess teaching quality and student outcomes in UK higher education providers by rating them as Gold, Silver or Bronze. TEF year two ratings were awarded in 2017 and were judged based on benchmarked metrics - considering the context of each provider and differences in students' backgrounds, entry qualifications and subjects studied - and a qualitative provider submission. Both the data and qualitative submission were considered by a panel of experts before the final judgement was made⁴.

National Student Survey

The NSS is mainly aimed at final-year undergraduate students and covers a number of aspects of the student experience including 'overall satisfaction'. To date, nearly three million students have taken the NSS and return rates are consistently around 70 per cent. In 2017, students at a number of alternative providers participated in the survey for the first time⁵. NSS performance can be assessed by absolute performance - what was your score - or by benchmarked performance - how far did your score differ from what would be expected considering the courses you teach, the characteristics of your student population, and so on. Both the absolute and benchmarked performance of providers have been considered as part of this research⁶.

Results

The overall finding of this research is that aggregated student reviews are positively associated with providers' subsequent APR, TEF and NSS outcomes. This is true of each individual data source, each is positively associated with the outcomes of the APR, TEF and NSS. But the collective-judgement score – created by combining all the reviews from every data source in a fixed period - proves an even more effective predictor than the individual data sources for APR and TEF outcomes, whereas the opposite is true for the NSS.

A poor collective-judgement score does not guarantee a provider will perform poorly on other quality measures; however, a provider with a poor collective-judgement score has a greater probability of performing poorly on other quality measures. Likewise, a provider with a good collective-judgement score is not guaranteed to be performing well on other quality measures but has a greater likelihood of doing so.

The relationship between collective-judgement scores and other measures of quality

APR outcomes

Figure 3 is a box plot of the collective-judgement scores for providers that were judged as either 'meets requirements with an action plan' or 'meets requirements'. The mean and median collective-judgement score prior to the APR data submission deadline is lower for the 14 providers judged 'meets requirements with an action plan' with a robust collective-judgement score available, than for the 173 providers judged 'meets requirements' with a robust collective-judgement score available.

Some caution must be taken with these results given the limited number of providers judged 'meets requirements with an action plan'; however, it is encouraging to see that there is limited overlap in the interquartile ranges, suggesting differentiation between providers' subsequent APR outcomes based on their collective-judgement scores.

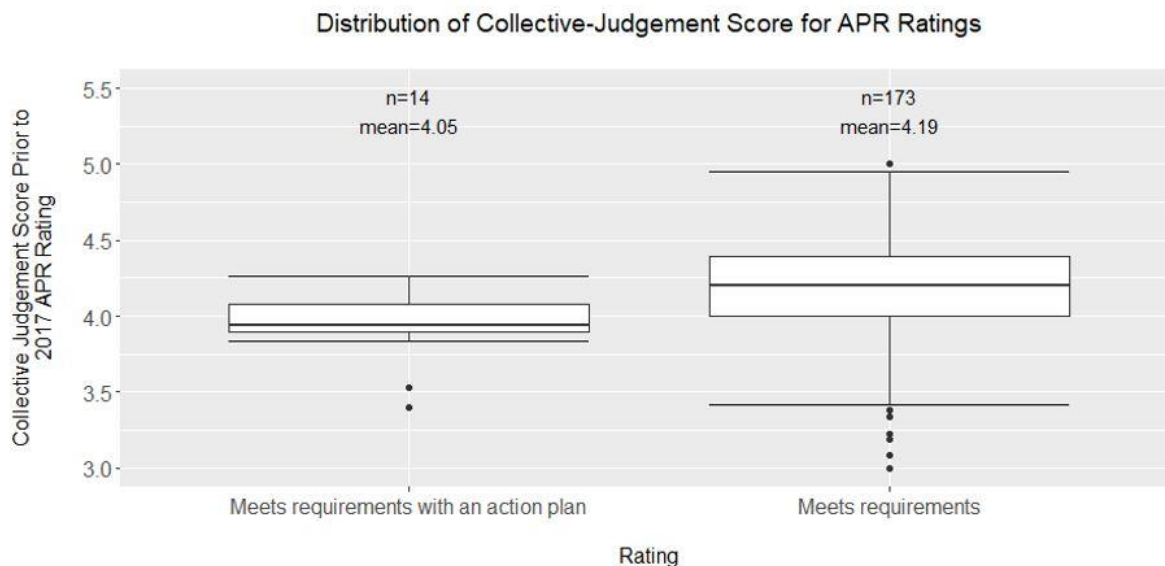


Figure 3: A boxplot of the collective-judgement scores on the data submission deadline for the 2016-17 APR.

TEF ratings

Figure 4 shows that the relationship between the collective-judgement scores and the TEF outcomes is similar to the relationship between the collective-judgement scores and APR outcomes. Both the mean and median collective-judgement score for Bronze-rated providers are lower than the mean and median

collective-judgement scores for Silver-rated providers, which in turn are lower than the mean and median collective-judgement scores for Gold-rated providers.

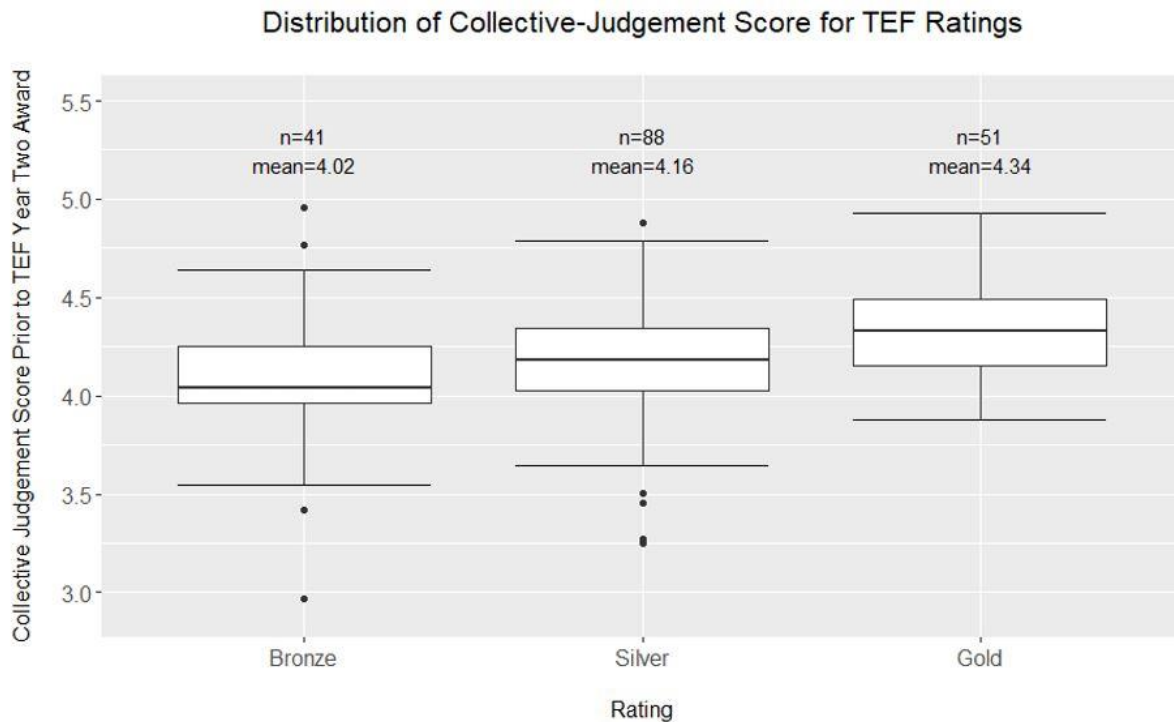


Figure 4: A boxplot of the collective-judgement scores on 22nd April 2017, 60 days prior to the announcement of the TEF awards.

For the TEF analysis, there is greater overlap between the interquartile ranges and the whiskers of the box plots for the Bronze, Silver and Gold-rated providers. This suggests that, whilst the collective-judgement score provides a statistically-significant predictor of a provider's TEF outcome overall, there is a significant variation of scores within the Bronze, Silver and Gold-rated providers and a number of outliers that do not follow the general trend. It is interesting to note that the distribution of the collective-judgements scores was greater for Gold providers than for Bronze or Silver providers. This suggests that it may be more challenging for students to identify Gold provision than Silver or Bronze provision.

NSS overall satisfaction scores

The analysis of the relationship between student reviews and the outcome of the NSS is necessarily different from the analysis for the APR and TEF outcomes, as the NSS result for each provider is a continuous score out of 100, rather than a discrete rating. Instead, the relationship between the collective-judgement score and NSS outcomes is assessed by the correlation between the two variables using the Pearson correlation coefficient.

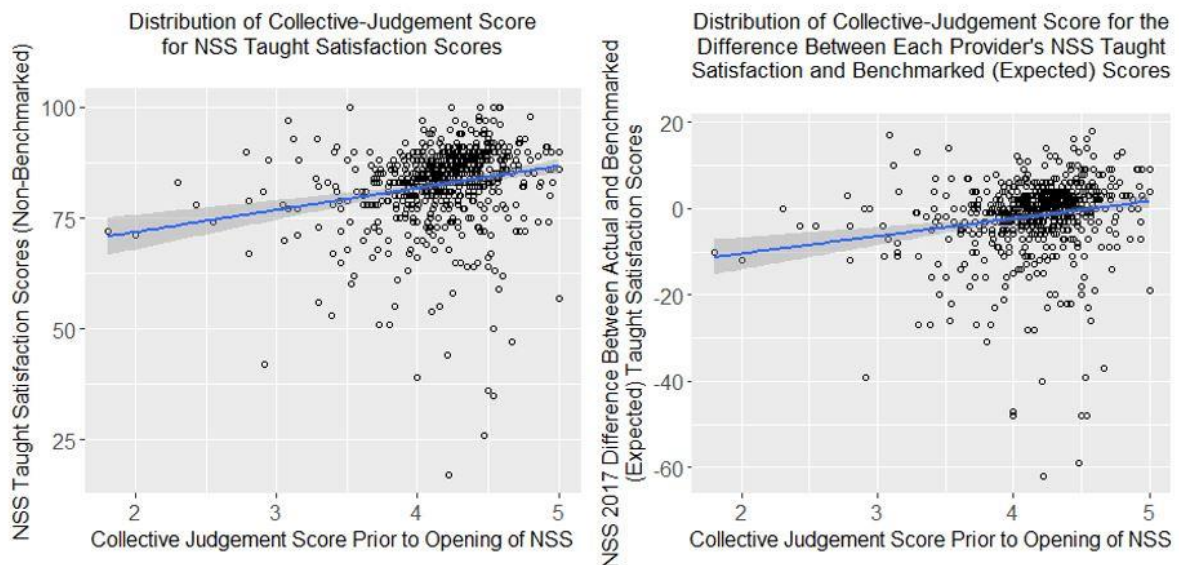


Figure 5: Two scatterplots of the collective-judgement scores on the opening day of the 2015, 2016 and 2017 NSS and, respectively, the non-benchmarked and benchmarked NSS overall satisfaction scores for taught students. Both plots feature a linear line of best fit with confidence intervals shown by the dark grey banding around the line.

Figure 5 shows there is a weak positive correlation between the collective-judgement score and both the non-benchmarked and benchmarked 'overall satisfaction' score (Pearson's correlation coefficients of 0.21 and 0.18 respectively). This suggests that, the higher a provider's collective-judgement score on the opening day of the NSS, the greater the likelihood that they will do well on the NSS. This is far from a robust relationship with a significant number of providers doing well on the NSS with a poor collective-judgement score and vice versa.

Discussion

This research is a first examination of whether, as has been shown to be the case in healthcare, aggregated user feedback might be able to identify the quality of higher education provision. The association between the collective judgement of students and existing quality measures suggests that it can. Moreover, it can do so in a more timely and less burdensome way than existing measures. The use of collective judgement however is reliant on a sufficient volume of reviews that, at present, is generally only available for large and medium sized providers.

Our results suggest that unsolicited student reviews are advantageous for measuring public perception and response, and therefore may be of particular utility for regulatory bodies who currently rely on more traditional surveys and metrics, and also have a new steer to act as a student champion^{7 8}. Whilst no one would suggest monitoring unsolicited student reviews should supersede other methods of oversight, such an approach offers several advantages over existing methods and may be a powerful new regulatory tool. Advantages include:

- real time feedback rather than reflective judgements at the end of a module/year
- feedback from more than final year undergraduates or recent leavers
- feedback based on what is important to students
- the ability to explore sector-wide issues, such as free speech, contact hours, or essay mills
- organisational learning to improve quality
- peer comparison
- prioritisation of quality interventions



There is also more that can be done with the data available. For example, time series can be developed to predict future movements accounting for seasonality and other factors, and the tool could focus on specific dimensions of quality such as teaching and learning or pastoral care. Alongside these current and future benefits there are challenges that must be considered when looking to make use of student reviews for regulatory purposes.

The results also suggest that monitoring student reviews may provide some benefit in identifying and preventing declining quality before the consequences become too great. There are however a number of points to consider:

- need for interpretation and contextualisation of any sudden shifts in the collective-judgement score before any action is taken. Nuances in the data and the phenomenon of students focusing their attention on specific issues such as a provider's not divesting from fossil fuels, means that any sharp changes in collective-judgement scores must be contextualised.
- agreement needs to be reached over what is not deemed to be a relevant review, and additional action taken to implement the identification and removal of such comments.
- some reviews are not accompanied by text explaining the reason for the review score. Making it difficult to know whether the review relates to broad quality concerns or a specific issue, relevant or otherwise.
- impact of specific issues affecting scores is far more common in smaller providers due to smaller student numbers and therefore fewer reviews.

In this context, the tool offers an opportunity to broaden the current set of quality assurance mechanisms, both for regulators and providers capturing information on a range of activities related to organisational performance and allow more timely responses.

Metrics aimed at improving the quality of higher education have traditionally failed to capture the student experience in real time. The use of student voice can promote organisational learning; identify system strengths and weaknesses and can be used to help prioritise or measure the efficacy of quality interventions within specific providers or sector wide. We have seen that student feedback has the potential to support regulatory oversight and organisational improvement and the specific challenges associated with using the data in this way, but there remain two wider issues for consideration.

First, some may question how reassuring it is that the collective judgement of students is positively associated with other quality measures? The APR, TEF and NSS processes are important to providers and the focus of significant attention, but they are not without their critics. For example, the TEF measures educational outcomes, but it does not directly measure teaching excellence. Is there a danger that a new measure would just be replicating the same issues present in existing metrics?

Second, are the privacy and ethical concerns that arise from monitoring student feedback⁹. For this research we have only used reviews that are publicly available. Specifically, for the Facebook data, we have only used reviews posted publicly on providers' review pages. We have not examined general posts made by students, even those posted on any other part of a provider's page, nor have we gathered any information about the users. At the very least, future research and any operationalisation of student feedback, must maintain strict ethical and privacy rules in order to be accepted.

Conclusion and future research questions

This research has explored whether the use of online student reviews can provide insight into the quality of higher education provision. After considering over 200,000 reviews and calculating a collective-judgement score from them, we found a positive association between the collective judgement of students and APR, TEF and NSS outcomes. The 'wisdom of crowds' phenomenon seems to hold true in this instance. Therefore, the use of the collective judgement of students could have significant benefits for regulators, providers and students. The volume of student feedback will continue to grow, and, the more it does, the more accurate the algorithms developed to classify the data and derive meaning from it will become.

This study represents only an introduction to the topic and a beginning of research into the use of the collective judgement of students. QAA and Alex Griffiths are working together on a pilot scheme with 10



providers from autumn 2018. The pilot is not part of any formal quality assessment but is testing how this pattern holds over time and assessing student, staff and organisational leaders' willingness to accept student feedback data, of the kind used here, as a valuable and helpful tool for broadening the scope of quality assurance and improving quality, the student experience and organisational learning. We hope that providers may also find value in this sort of analysis for their own internal monitoring, spotting the early warning signs and putting right issues before they warrant regulatory intervention.

Discussion questions:

1. What are the further policy and practice implications of this research for the sector?
2. At a time when people are increasingly careful about how their data is used by technology companies, would students accept their feedback being used by providers and regulators?
3. How do you think providers might engage with this source of student feedback?

Please submit your proposal by sending this form, in Word format, by 24 July 2018 to QAForum@eua.eu. The file should be named using the last names of the authors, e.g. Smith_Jones.doc. Please do not send a hard copy or a PDF file.

References:

1. Surowiecki J. The wisdom of crowds. London: Little, Brown 2004.
2. Griffiths A, Leaver MP. Wisdom of patients: predicting the quality of care using aggregated patient feedback. *BMJ Qual Saf* 2018;27(2):110-18.
3. HEFCE. Revised operating model for quality assessment: HEFCE; 2016 [Available from: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2016/201603/HEFCE2016_03.pdf accessed 28/04/2016 2016.
4. OFS. What is the TEF? 2018 [Available from: <https://www.officeforstudents.org.uk/advice-and-guidance/teaching/what-is-the-tef/>.
5. NSS. The National Student Survey 2018 [Available from: <http://www.thestudentsurvey.com/about.php> accessed 01/02/2018.
6. HEFCE. National Student Survey 2017 [Available from: <http://www.hefce.ac.uk/lt/nss/fag/#Q2> accessed 17/02/2018.
7. Gyimah S. Strategic Guidance to the Office for Students – Priorities for Financial Year 2018/19 - Letter from Sam Gyimah, Minister of State for Universities, Science, Research and Innovation, to Sir Michael Barber, Chair of the Office for Students 2018 [Available from: <https://www.officeforstudents.org.uk/media/1111/strategicguidancetotheofs.pdf> accessed 03/03/2018.
8. Higher Education and Research Act 2017.
9. Cadwalladr C, Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Observer* 2018 Sat 17 Mar 2018.