The Wisdom of Students:

Monitoring Quality through Student Reviews

Updated analysis

Alex Griffiths

July 2019

Contents

Introduction

Rapid digital change over the past 20 years has made sharing our opinions and finding those of others easier than ever before. Many first-time undergraduate students will have grown-up in a world where social media and customer reviews are a part of everyday life.

In 2018, we set out to explore whether the wealth of individual online feedback about higher education providers represented the discrete and disconnected views of individuals with extreme experiences (giving limited insight into wider patterns in the quality of provision). Or whether, taken as a whole, they could provide more obvious pointers towards changes that might improve services.

The answer we found was a positive association between the ratings aggregated from individual data sources, specifically Facebook, Whatuni.com and StudentCrowd.com, and TEF, APR and NSS outcomes. QAA published the findings in The Wisdom of Students.

Moreover, when the reviews from these data sources were combined we saw a 'wisdom of crowds' effect - the greater number of reviews capturing a greater range of experiences across different platforms resulted in an even stronger positive association covering a greater number of providers. The better the reviews for a provider in the year prior to other quality measures being assessed, the more likely it was to do well on those quality measures, and vice versa. However, probabilities are not certainties. While a provider reviewed negatively online was likely to perform poorly on other quality measures, it was not a given.

The early results were promising but just the beginning. Would the positive association hold over time, or was it a singular, chance relationship? Would including additional data sources, such as Twitter, enhance the 'wisdom of crowds' effect? What themes and patterns could we identify from the wealth of feedback we had available? This follow-up report begins to answer these questions.

Data

To determine whether the positive association between aggregated reviews and other quality measures held, and whether additional data sources added further value, we repeated the test from our earlier report including the more recent (2018) TEF and NSS results.

The average review scores - all measured on a scale of one to five stars with five being the best - were calculated for each data source over the 365 days prior to TEF and NSS results being determined. The relationship between these average scores and TEF/NSS outcomes was then assessed for all providers where there had been five or more relevant reviews over the 365 days.[1]

The majority of reviews were relevant: however, some quality control was required. For example, knowing that the reviews page of universities will be visited by current and prospective students, enterprising students use their provider's Facebook reviews page to advertise books for sale or rooms for rent, rather than provide a genuine review. We developed a custom algorithm to identify and remove irrelevant reviews.

---

[1] For a full explanation of why a 365-day moving average, rather than, for example, a 180-day moving average, please see the earlier report: www.qaa.ac.uk/docs/qaa/about-us/the-wisdom-of-students-monitoring-quality-through-student-reviews.pdf

Reviews across each and of the data sources were overwhelmingly positive averaging over four out of five stars (with the exception of StudentCrowd.com reviews), and providing a positive reflection of UK higher education provision. The data sources were as follows:

Whatuni.com

With over 120,000, Whatuni.com had the greatest number of reviews. The number of new reviews has fallen significantly in recent years with fewer than 13,500 in 2018 compared to more than 31,000 the year before. As the name would suggest, reviews focused heavily on universities.

Facebook

Over 53,000 relevant reviews, and some 2,000 irrelevant reviews, are available on providers' main Facebook pages. Facebook has good coverage of all provider types although, in common with Whatuni.com, has seen a decrease in the number of reviews in 2018. To be clear, Facebook comments were not taken from individual's Facebook pages, only the public 'reviews' page of a provider's main Facebook page where it is explicitly public.

StudentCrowd.com

StudentCrowd.com publicly detailed over 13,500 provider and course-level reviews, including over 2,000 in 2018.

StudentHut.com

StudentHut.com featured over 9,000 publicly-accessible, provider-level reviews focusing on universities. As a newer review site, nearly all of these reviews were posted in 2018 so a full 365 days' worth of reviews was not available to assess the relationship between StudentHut.com reviews and TEF and NSS scores in isolation. But the reviews still contributed towards the aggregated 'Collective Judgement Score' (CJS) for a provider.

Twitter

Twitter differs from the other data sources in that comments are not scored by users on a one-to-five star scale, nor are most comments relevant. To overcome this issue, we developed custom models to classify who made the comment (to exclude comments from providers and their staff), the contents of the comment and, if relevant, its sentiment on a one-to-five star scale.

As with Facebook reviews, we had to be careful to balance utility with user privacy. We only gathered tweets where the user had actively 'mentioned' a provider, that is included a provider's main account in the body of the tweet making the provider aware of the comment. Tweets simply naming a provider, or using a hashtag relating to that provider, have not been gathered. From our discussions with providers, this appears to be a level of privacy greater than many of their own social media teams adhere to.

Only gathering tweets mentioning providers resulted in 6.25 million tweets being gathered between September 2017 and January 2019. Of these, 2.25 million were original, that is neither a retweet nor created by a provider itself.

Google

Since the start of 2017, there have been just under 17,000 reviews of UK higher education providers on Google, of which over 15,500 are relevant. The ability to leave a Google review for any business or organisation means that nearly all providers, including further education colleges and alternative providers, had some reviews.

Results

Does the association between the data and TEF and NSS outcomes still hold, and is the same true of new data sources?

TEF Ratings

Of the 87 non-provisional TEF ratings made in April 2018, the majority were to smaller providers that had not previously received a rating. With fewer students studying at smaller providers, it is harder to obtain enough comments from a single data source in 365 days to make a robust judgment than it is for a larger provider.

Table 1 below shows that the relationship between the individual data sources already assessed and TEF ratings does hold for Whatuni.com and StudentCrowd.com in 2018, albeit based on fewer providers than would be ideal for a robust assessment. Facebook reviews still distinguish between Gold and non-Gold providers. However, where there were sufficient numbers, Facebook reviews were marginally more positive for Bronze providers than Silver providers.

Data Source	Metric	TEF Rating		
		Bronze	**Silver**	**Gold**
Whatuni.com	Providers covered	4	9	8
	Number of ratings	560	2758	2377
	Average CJS	4.01	4.09	4.26
Facebook	Providers covered	9	16	6
	Number of ratings	194	490	248
	Average CJS	4.11	4.01	4.57
StudentCrowd.com	Providers covered	2	8	4
	Number of ratings	31	72	86
	Average CJS	1.81	2.82	3.89

Table 1: The relationship between previously assessed data sources and the TEF Year 3 (2018) outcomes where there have been ≥ 5 reviews in the preceding 365 days.

For the three new data sources - Google, Twitter and StudentHut.com - only Google reviews were (a) available and (b) could be obtained in an economically-viable way far enough back to permit a 365-day moving average prior to the 2018 ratings being awarded.

Data Source	Metric	TEF Award		
		Bronze	Silver	Gold
Google	Providers covered	8	21	5
	Number of ratings	219	1,005	325
	Average CJS	3.88	4.02	4.40

Table 2: The relationship between Google reviews and the TEF Year 3 (2018) outcomes where there have been ≥ 5 reviews in the preceding 365 days.

Google reviews mirror the same positive association we have seen with other data sources previously: the average Collective-Judgement Score for a Bronze-rated provider is less than that of a Silver-rated provider, which is in turn less than that of a Gold-rated provider.

Data Source	Metric	TEF Award		
		Bronze	Silver	Gold
All data sources combined	Providers covered	14	28	11
	Number of ratings	1,080	4,723	3, 268
	Average CJS	3.88	4.03	4.41

Table 3: The relationship between all reviews combined and the TEF Year 3 (2018) outcomes where there have been ≥ 5 reviews in total in the preceding 365 days.

When all data sources are combined, including relevant comments from Twitter and StudentHut.com, we see greater coverage (as we would expect) and the positive association becomes clearer and more distinct.

NSS Outcomes

More providers participating in the NSS than receiving new or updated TEF ratings in 2018 means that far more providers met the five-review threshold for a data source to be assessed.

Data Source	Correlation with NSS Score - 2018	Correlation with Benchmarked NSS Score - 2018	Providers Covered
Whatuni.com	0.17	0.19	129
Facebook	0.21	0.25	153
StudentCrowd.com	0.19	0.13	81

Table 4: The correlation between previously assessed data sources and the 2018 NSS 'Overall Satisfaction' results where there have been ≥ 5 reviews in the preceding 365 days.

As with previous individual years, all data sources show a weak, positive correlation with both the NSS non-benchmarked and benchmarked outcome.

Data Source	Correlation with NSS Score - 2018	Correlation with Benchmarked NSS Score - 2018	Providers Covered
Google	0.17	0.16	180
All data sources combined	0.11	0.13	308

Table 5: The correlation between Google reviews and all reviews combined and the 2018 NSS 'Overall Satisfaction' results where there have been ≥ 5 reviews in the preceding 365 days.

In line with the existing data sources, Google reviews and the combined reviews have a weak positive correlation with little difference between the non-benchmarked and benchmarked performance. It is clear there is some relationship between each data source and both the non-benchmarked and benchmarked findings for the 'overall satisfaction' question of the NSS. However, this positive correlation is far from strong. Plus, when the comments are aggregated the correlation weakens, suggesting they have different underlying relationships.

Gathering reviews systematically across multiple platforms, filtering out those reviews and comments that are irrelevant, and categorising and scoring those that remain (where necessary) is an extremely challenging and constantly evolving process. The results above suggest that there is value in persevering with this approach. Student reviews are an effective predictor of TEF ratings and appear to loosely align with NSS overall satisfaction scores.

Being perfectly aligned or otherwise with distinct measures of quality does not, of course, indicate the value of student reviews. Both the NSS and TEF have come in for significant criticism (see for example Ashwin (2018); Cheng and Marsh (2010); Fielding et al (2010); Williams (2015) and Yorke (2009)). Rather, it serves to provide some re-assurance they are not an isolated collection of grievances.

Conclusion

The findings here re-affirm our previous findings, extend to new data sources, and suggest that the reviews have value in their own right. Indeed, the regulatory monitoring of user reviews, while not advocated by QAA, has already been proven in the English healthcare system (Griffiths and Leaver, 2018).

References

Ashwin, P (2018) TEF 2018: Can students trust the gold standard? Times Higher Education. 6 Jun 2018 [online]
www.timeshighereducation.com/blog/tef-2018-can-students-trust-gold-standard

Cheng, J H S and Marsh, H W (2010) 'National Student Survey: Are differences between universities and courses reliable and meaningful?' Oxford Review of Education, 36, 693-712.

Fielding, A, Dunleavy, P J & Langan A M (2010) Interpreting context to the UK's National Student (Satisfaction) Survey data for science subjects, Journal of Further and Higher Education, 34:3, 347-368.

Griffiths A and Leaver M P (2018) Wisdom of patients: predicting the quality of care using aggregated patient feedback, BMJ Quality & Safety, 27(2), pp 110-18.

Williams, J (2015) The National Student Survey should be abolished before it does any more harm. The Guardian. 13 Aug 2015 [online]
www.theguardian.com/higher-education-network/2015/aug/13/the-national-student-survey-should-be-abolished-before-it-does-any-more-harm.

Yorke, M (2009) 'Student experience' surveys: Some methodological considerations and an empirical investigation'. Assessment and Evaluation in Higher Education, 34, 721-739.