Assessing the quality of UK medical schools: what is the validity of student satisfaction ratings as an outcome measure?

Tim Lancaster, Director of Clinical Studies, Oxford University Medical School

Thomas R. Fanshawe, Statistician, Nuffield Department of Primary Care Health Sciences, University of Oxford
Background

In 2015, the UK government announced its intention to implement a Teaching Excellence Framework (TEF) for universities, with results possibly influencing funding decisions in a similar manner to the Research Excellence Framework (Grove 2015). There is as yet no detail of what metrics will be used in the TEF. The National Student Survey, which assesses student satisfaction with course provision, has been offered to all final year students at Higher Education Institutions since 2005. NSS scores are readily available and are already used as a quality measure by medical schools and their potential applicants: it would be logistically straightforward to use these scores as a TEF metric (Grove 2015).

The value of the NSS as an outcome measure is contested. One recent commentator argued that ‘this survey and its outcomes should undoubtedly form an important part of the TEF; …it should be heavily weighted in comparison to other metrics, too’ (Owen 2015). Others have suggested that the NSS may be associated with negative educational outcomes, such as over simplification of curriculum material, lower passing standards and grade inflation: ‘an intellectual race to the bottom’ (Williams 2015). This lack of consensus arises partly because of the difficulty, for many university courses, of defining measurable outcomes that could form a reference standard for evaluating the NSS.

The vocational nature of university courses in medicine has a clear conceptual outcome: all medical schools set the objective of graduating effective doctors. Metrics for measuring achievement of this objective have historically been lacking. In 2015, the General Medical Council (GMC) for the first time co-ordinated the publication of comprehensive data on outcomes of UK medical graduates in postgraduate training by Primary Medical Qualification (medical school of graduation). These included graduate self-report of
preparedness for practice, pass rates in postgraduate examinations and outcomes of annual progress reviews. Postgraduate examination performance and annual progress reviews are objective measures relevant to performance as a doctor. To inform debate about metrics appropriate for a TEF in medicine, we aimed to assess the extent to which NSS scores correlated with the reference standard of these performance based outcomes.

**Methods**

The National Student Survey consists of 22 questions to which students are asked to respond on a five point Likert scale from definitely agree (5) to definitely disagree (1). The 22 questions cover seven domains: Teaching; Assessment and Feedback; Academic Support; Organisation and Management; Learning Resources; Personal Development and Overall Satisfaction. We calculated medical school NSS scores as the mean score out of 5 across the 22 questions. To moderate year to year variations, and because postgraduate examination performance includes data from graduates of different years, we took the mean across six years from 2008 to 2014. The score for postgraduate examination performance was the percentage pass rate across all postgraduate examinations by Primary Medical Qualification (PMQ), as provided by the GMC for 2014 (www.Gmc-uk.org/education/25496.asp). We also assessed the relationship between examination performance and qualifications at entry to medical school, using the average UCAS tariff of each medical school from 2008 to 2014 based on data from the Higher Education Statistics Agency, as reported by the Complete University Guide. We calculated Pearson correlation coefficients for each of these relationships, and partial correlation coefficients to assess the relationship between two of the variables, controlling for the third. We calculated the correlation between postgraduate examination performance and on the job performance as judged by number of unsatisfactory
outcomes in annual progress reviews. Deaneries and local education and training boards review the progress of all doctors in training on at least an annual basis, and award them with either an Annual Review of Competence Progression (ARCP) or Record of In-Training Assessment (RITA) outcome, depending on when they started their training. As examination failure may lead to an unsatisfactory ARCP/RITA outcome, we used outcomes that excluded examination performance. We grouped schools according to the number of specialties in which ARCP/RITA outcomes were statistical outliers, significantly above or below the national average for unsatisfactory outcomes. In additional analysis, we assessed the correlation between examination performance and rankings of medical schools in broadsheet subject tables, which use multiple putative quality metrics. We calculated Spearman rank correlation coefficients for performance in postgraduate examinations with rank in the medicine subject tables of the Complete University Guide, the Guardian University Guide and the Times/Sunday Times University Guide.

**Results**

Mean NSS scores ranged from 3.61 to 4.59 (mean 4.00). Pass rate in postgraduate examinations ranged from 50% to 87% (median 71%). Numbers of outliers for ARCP/RITA ranged from -3 to +3 (median -1).

NSS score did not correlate with postgraduate examination performance, $R = 0.17$ [95% confidence interval -0.21, 0.51], $p=0.38$, Figure 1. One school was ranked first on both measures, thirteen schools were ranked higher by examination performance than by NSS score, and 14 were ranked lower.
By contrast, there was a positive correlation between average entry qualifications and performance in postgraduate examinations, $R = 0.73$ [95% confidence interval 0.48, 0.87], $p<0.001$, Figure 2.
The correlations between NSS score and postgraduate examination performance, controlling for average entry qualifications, and between average entry qualifications and postgraduate examination performance, controlling for NSS score, did not greatly change (partial $R = 0.11$ and 0.67 respectively).

Figure 2. Entry qualifications (UCAS tariff) and percentage pass rate in postgraduate examinations
There was a positive correlation between postgraduate examination performance and ARCP outcomes, $R = 0.48$ [95% confidence interval 0.10, 0.74], $p=0.02$. Broadsheet league table positions calculated on a range of metrics correlated positively with postgraduate examination performance with the best correlation achieved by the Complete University Guide, $ρ = 0.60$, $p<0.001$. Significant but less strong correlations were found for the Times, $ρ = 0.48$, $p=0.01$ and Guardian University Guides, $ρ = 0.38$, $p=0.05$.

**Discussion**

Our results show that it is possible to achieve high student satisfaction ratings and high graduate performance; in our analysis, the highest ranking school for examination performance also had the highest NSS score. However, for most schools there was a lack of correlation between the two measures, and in some cases the discrepancies were large (Figure 1). This finding raises concerns about validity of the NSS as a measure of quality for medical degree courses.

By contrast, while it did not explain all the variance, examination performance correlated significantly with the qualifications of students at entry to medical school. This is consistent with the findings of previous analyses of the MRCP examination (McManus 2008). The ability to attract better students may be a reasonable, market-based, marker of quality, but this correlation gives no indication of the value added by individual medical courses. Examination performance corrected for entrance qualifications might be a better marker of medical school quality.

As with a previous analysis (Wakeford 2015), we found inconsistent correlation of broadsheet league tables with rank based on postgraduate examination performance. These
tables, based on panels of metrics including NSS scores and entry qualifications, are superior to the NSS alone as quality markers. However, none perform as well as entrance qualifications alone in predicting outcome. Further research is needed to determine if there is a combination of metrics with better correlation to outcome.

There are a number of limitations in the data used for this analysis. GMC data on examination performance were not available for all London medical schools because some examinees elected to be coded as graduates of the University of London. UCAS tariffs were not available for graduate entry courses. For some newer medical schools, the number of graduates entered for postgraduate examinations in 2014 is relatively small and the estimates of success correspondingly less precise. Data on all postgraduate examination performance is available for only one year, 2014, and includes data from over 100 different examinations with different formats. Examination results by PMQ may be confounded by postgraduate educational experiences. Nevertheless, previous analyses of the MRCP and MRCGP examinations suggest that correlations between examination performance and medical school of origin are consistent over time (McManus 2008, McManus 2013).

The objective of medical training is to produce high quality doctors, so it is legitimate to ask whether examination performance is a valid measure of on the job ability. We found a correlation with ARCP/RITA outcomes which provides some evidence that examination performance may indeed be a useful proxy for clinical performance.

Lack of correlation with outcome does not necessarily make the NSS an irrelevant metric. NSS scores could, for example, reflect the extent to which students feel taught, cared for, supported and listened to by their medical school. As such they may represent one type of
marker of the quality of an educational environment. However, student satisfaction as a measure of the quality of education is limited by the restricted ability of respondents to compare the course they have experienced with those they have not. Respondents may also find it difficult to recognise which features of a medical course will be of most help to them in postgraduate training. It is a plausible hypothesis that there are features of some medical courses that are difficult and unpopular with students, but lead to more capable graduates. There is, for example, evidence of variation between UK medical schools in the volume and intensity of summative assessments (McManus I.C. personal communication). Although our study does not address this question, tougher assessments could plausibly be negatively correlated with student satisfaction and positively correlated with later examination success. Differences in curriculum content such as a greater or lesser emphasis on biomedical science might similarly be relevant. Defining which, if any, features of medical courses affect variation in performance following graduation is an important area for future research.

Conclusion

Our results show that levels of student satisfaction in different UK medical schools, as measured by the National Student Survey, do not correlate with the performance of their graduates in postgraduate assessments. Assuming that the overriding measure of the success of a medical school is its ability to graduate competent doctors, the NSS appears to have little or no value as a quality metric for the Teaching Excellence Framework in medicine. Examination performance, which correlates with on the job performance, is a better measure of outcome in medical education. However, the effects of differences in ability of student cohorts on performance must be taken into account in assessing the value added by medical schools. Outcome based metrics are more difficult to define for those university courses (the
majority) that lack a national reference standard. Our findings suggest that caution is required in using NSS results as a proxy measure of university course quality.

References

Owen D. How to build the teaching excellence framework. *Times Higher Education* June 15, 2015

Grove J. How might a teaching excellence framework be built? *Times Higher Education* June 23 2015


McManus IC, Wakeford R. PLAB and UK graduates’ performance on MRCP(UK) and MRCGP examinations: data linkage study. *BMJ* 2014;348:g2621.

Wakeford R. Medical schools should stop cherry picking league table data. *Student BMJ* 2015;23:h1359

Williams J. The National Student Survey should be abolished before it does any more harm. *The Guardian* 13 August, 2015

Acknowledgments

Chris McManus, Jonathan Price, Helena McNally, William James, Damion Young, and Sasha Shepperd provided helpful comments on earlier drafts of this manuscript