Exploring the value-added of postgraduate medical education: A preliminary evaluation

Journal:	*Medical Education*
Manuscript ID:	MED-2015-0073
Manuscript Type:	Research Papers
Keywords:	Postgraduate training, Testing/Assessment, Quantitative Research Methods

Exploring the value-added of postgraduate medical education: A preliminary evaluation

Abstract: Context: Increasing pressure is being placed on external accountability and cost efficiency in healthcare education internationally. Robust cost analyses using evidence-based approaches are required to inform stakeholders about the utility and value of education. We present illustrative data of the value-added of postgraduate medical training interventions and debate whether models of value-added used in other education settings are appropriate for medicine.

Method: We analysed anonymised, historical selection (entry) and licensure (exit) examination results for trainees sitting the UK MRCGP licensing examination in 2010. Entry data were assessment scores from the UK GP selection process comprising; (1) a clinical problem-solving test; (2) a situational judgement test; and (3) a selection centre. Exit data was an applied knowledge test (AKT) in the licensing exam. Exit data were matched to the selection results anonymously. Ordinary least squares regression analyses were used to model differences in attainment in the exit examination based on performance in entry assessments (the value-added score). Our results were aggregated up to the regional level for comparisons.

Results: We discovered significant differences in value-added scores between regional training providers. Results suggest three regional training providers confer significant value-added, based on the prior attainment of the trainees. By contrast, results for another regional training provider were significantly lower than predicted based on the attainment of trainees in entry examinations.

Conclusions. Our illustrative data analysis offers direct evidence of the potential utility of

value-added analysis in postgraduate medical education and training. We present evidence of differential improvements in trainee's attainment at regional level in a licensure examination, independent from the level of attainment at entry to training. This value-added analysis method is complex and controversial and has important limitations. Results indicate the method offers important insights that may be used to drive continuous improvement in postgraduate medical education in future.

Introduction

Medical education is expensive[1]. Internationally, significantly increased emphasis is being placed on external accountability and cost efficiency in education and training within healthcare. Providers are under mounting pressures from trainees, regulators and policymakers to demonstrate and improve the quality and effectiveness of medical education and training provision[2,3]. Robust cost analyses using evidence-based approaches are needed to inform funders and users about the utility and value of medical education interventions. This will aid more accurate and effective commissioning in future[4]. Such approaches are already adopted to evaluate the value-added by such interventions in other fields of education. Should these approaches be used to evaluate value-added in medical education interventions? Here we highlight this issue to encourage debate in the academic and practitioner communities. We explore key concepts in value-added measurement and examine its relevance for postgraduate medical education and training by presenting results from an illustrative data analysis.

Using historical data, this study considers the appropriateness of models to explore potential future development of value-added measures in the field of medical education. Our intention is to consider whether this type of analysis could usefully contribute to identification of models of good practice in training interventions, and to learn about the effectiveness of training models. Specifically, we address two key research questions:

(1) *Is it possible to devise a method for assessing value-added in postgraduate medical education and training interventions of different types, in different regions, and with different profiles, and does the method yield more valid and accurate conclusions than current performance indicators?*

(2) *What are the implications of implementing a system of evaluating value-added? Specifically, is it appropriate and desirable to assess the learning outcomes of different education and training providers to inform future commissioning activities*?

Drivers for assessing value-added in medical education and training

Unlike postgraduate medical education, assessment of value-added in higher education has been debated in the public domain for many years, with a view towards enabling applicants to make more informed judgements about a medical school before applying[5,6]. In the higher education context, assessing value-added is driven by an increasing trend to provide potential students with a variety of information on courses. It could be argued that this will allow potential students to make more cognisant choices. Similarly, institutions are keen to use value-added indicators as a vehicle for marketing in an increasingly competitive environment. Moreover, funders of education (whether governments, institutions or individuals) are equally keen to ensure maximum return on their investment.

There is a wealth of research on value-added arising from evaluations of education in schools. Some researchers claim that race, socioeconomic status, class size, and classroom heterogeneity are poor predictors of student academic growth, and that the effectiveness of the teacher is the major determinant of student academic progress[7]. Other researchers have fiercely debated the impact, accuracy and appropriateness of value-added assessments[8], implying that such evaluations may over-simplify the role of educators and that results can be politically divisive.

In principle, it may be timely to consider an assessment of value-added of education and training interventions in postgraduate medical education. For example, in the UK in early 2012, the regulator, The General Medical Council, began a consultation on new proposals for the recognition and approval of medical educators to improve the quality and consistency of medical training in all regions, which led to the introduction of standards[9]. In these new proposals, regional education providers are responsible for demonstrating how local arrangements maintain high quality training standards. An important issue for providers is how best to make a fair and accurate appraisal of the relative quality and standard of postgraduate training across regions.

Key concepts in value-added measurement

Value-added models are used in many different settings to evaluate, monitor, and improve an education provider and/or other aspects of an education system[10,11]. The value-added approach stems from the assumption that educators can add value to the achievement of their students. The approach is based on measuring student progress for one training provider, usually by tracking relative attainment outcomes over a given time period, in comparison to student progress achieved by other training providers. The concept has relevance to medical education as it can be applied to other quantifiable trainee outcomes, such as measures of vocational competence.

Several different value-added methods exist, but approaches usually involve estimating expected scores at the end of the education and training intervention for individuals, based on the average attainment of individuals from a representative sample or population with the same or similar prior attainment scores. The value-added score for a provider is then estimated, for example via the average difference between expected scores and actual achievement across all individuals at that provider. Positive value-added scores indicate student progress that is better than expected, while negative value-added scores indicate student progress that is worse than expected. Thus for each provider the value-added score represents a relative measure of the typical progress made by individuals between different stages of education. The most effective providers would be those in which student progress exceeds expectations, having also taken into account statistical uncertainty in measurement. Meaningful comparisons can then, in principle, be made between providers and across time. Regions that have more applicants and can be more selective will tend to have better raw outcomes. However, this may be due more to the quality of the intake than the quality of training provided. By taking into account prior achievement, it is possible to assess the relative amount of progress made by individuals independent of attainment levels at point of entry to the training programme. This may facilitate the identification of the most effective training interventions, allow best practice to be shared between providers, and ultimately improve training standards across the board.

Performance in licensing exams is an existing measure for assessing postgraduate training standards in medicine. However, raw examination results have considerable weaknesses when used as absolute measures of training standards. The most serious weakness is that the use of raw examinations results fails to take into account individuals' prior attainment. Conceptually, value-added measures were developed to account for prior attainment, and to estimate more fairly and accurately the additional knowledge gained during specific periods of training. In postgraduate medicine, *value-added* may relate to the level of trainee's progress and growth in knowledge, skills, abilities, and other attributes, gained over time as the result of a training programme, relative to the progress of trainee's on other equivalent training programmes. Worldwide, postgraduate medical trainees take standardised assessments both during and at the end of training to demonstrate competence as an independent practitioner. Examining profiles at selection using results from predictive validity studies and linking these with performance in licensing assessments at the end of training [10,11] (e.g. Membership of the UK Royal College of General Practitioners; MRCGP) provides an opportunity to assess value-added in medical education and training interventions, both within and across regions.

Method

We obtained anonymised, historical selection (entry) and exit examination results for a sample of trainee doctors sitting the UK MRCGP licensure examination in 2010. The entry results (input measures) were assessment data from the UK general practitioner (GP) specialty training selection process. Training in UK general practice typically lasts for three years, so the data at point of selection were matched with an applied knowledge test at the end of training as part of the MRCGP exam. The data were matched by the UK GP National Recruitment Office to protect individuals' anonymity.

The selection system is designed to process thousands of applicants every year, and targets six core attributes that were identified by a multi-method job analysis study[12,13] (*empathy*, *communication skills, problem-solving, professional integrity, coping with pressure, and clinical expertise*). The selection method comprises three elements. First, applicants sit two machine-marked tests: (i) a *clinical problem solving test* (CPST) comprising questions that require applicants to apply clinical knowledge to solve problems reflecting diagnostic processes or develop management strategies for patients; (ii) a *situational judgement test* (SJT) targeting non-academic attributes (e.g. empathy, integrity, coping with pressure) where applicants are presented with written depictions of professional dilemmas they may encounter at work and are asked to identify an appropriate response from a list of alternatives[14]; and (iii) a *selection centre* (SC) using job-relevant simulations (a patient consultation, and group and written simulation exercises) to target clinical and non-clinical attributes[12,15]. Typically, 10-20% of applicants are rejected at shortlisting, with a further 20-30% selected-out at the final stage selection centre. Good evidence of the predictive validity of the selection assessments has been demonstrated three months into training, after one year of training, and at the end of training[16, 17, 18].

The exit results (i.e. outcome measures) were results from the UK MRCGP licensing exam in addition to various workplace assessments during training. Sufficient data were available for one of the licensing assessments called the *Applied Knowledge Test* (AKT), which is a three-hour knowledge assessment comprising 200 items focusing on clinical medicine, critical appraisal and health informatics.

Anonymised outcome data (AKT exam results) were matched to the input data (selection results) by the GP National Recruitment Office. A total of seventeen deaneries (regional education and training providers) existed in the UK in 2010, and each region had varying numbers of trainees. To protect anonymity, we classified each deanery into either a small (n< 100 trainees recruited per annum), medium (n= between 100 and 150 trainees recruited per annum) or large (n> 150 trainees recruited per annum) training region. We then randomly selected nine of the 17 regions (three small, three medium and three large) to present illustrative findings, although the analyses were conducted using the complete dataset to ensure maximum rigour.

The input selection measures (CPST, SJT and SC) were entered into an ordinary least squares (OLS) multiple linear regression model. The outcome measure was the AKT exam of the MRCGP. Where trainees had taken the AKT exam more than once we used their average score. Outcome scores were standardised and are presented in standard deviation units.

Output from the regression analyses provided estimates of the value-added score (studentised residual) for each trainee. Individuals' scores were aggregated up to the regional level to compare training providers. The results for each region comprise the mean valued-added scores (studentised residuals) estimated by the regression analysis, as well as the associated average predicted score for trainees in the AKT exam and the average score actually achieved (n.b. only the value-added scores are presented in Table 1). Where the value-added scores are positive (i.e achieved scores in the AKT were higher than the predicted score at point of selection), the trainees were doing better than expected given their initial level of attainment. This may result from of the quality of the training they received. Conversely where an AKT value-added score is negative (i.e. achieved score is lower than predicted score), trainees were performing at a lower level than expected given their initial level of attainment.

Tests of statistical significance were conducted to indicate if each regional value-added (residual) score was statistically significant different from what would be expected (i.e. from zero). A one-way ANOVA was also employed to indicate simply if there was a statistically significant difference in the trainee's average value-added scores (residuals) between the nine regions presented.

Results

A sample of 2291 individuals from all 17 regions was available with matched selection data and AKT results. Figure 1 shows aggregated regional-level data from the nine randomly selected regions to illustrate how value-added was estimated. Each point represents one region. The regression line represents predicted AKT mean score, calculated on the basis of the mean baselines scores (CPST, SJT and SC) of the same trainees in 2007. Regions with points above the regression line have a positive value-added score (i.e. residual) suggesting that performance exceeded expectation. Regions below the regression line have negative value-added scores suggesting performance was below that expected. Information on the level of confidence associated with value-added scores (e.g. 95 per cent confidence intervals) is necessary to indicate whether discrepancies in performance of trainees in a region might have occurred by chance, or whether they represented a 'real' (i.e. statistically significant) difference.

INSERT FIGURE 1 HERE

Table 1 presents the results of the value-added analysis for each region for the AKT exam. When comparing performance at point of licensure (typically after three to four years of training) we observe significant differences in the value-added score between regional training providers.

A one-way analysis of variance comparing the residuals for the different training regions was statistically significant ($F(8,1128)=4.04$, $p<0.01$). This means that the differences in residual values between the training regions were greater than would be expected by chance. Two-tailed t-tests were used to explore whether the residuals for any training region were significantly different from zero (i.e. whether trainees in that region performed significantly better or worse than would be expected from their performance at recruitment). AKT results suggest that training regions 6 ($t=2.1$, $p<.05$), 9 ($t=3.0$, $p<.01$), and 16 ($t=2.2$, $p<.05$) confer significant value-added, given the profile of trainees at point of selection. By contrast, results for training region 10 (t=3.8, p<0.01) suggest that trainee achievement in the AKT was significantly lower than would be predicted given the trainee profile at point of selection.

It is important to view these findings as results of an illustrative analysis (to explore the methodology rather than draw conclusions about the regional training providers), and caution must be taken in interpreting the results due to the sample size and limited data available. In addition, if alternative outcome data was available for analysis using a different test modality (e.g. a clinical skills assessment as opposed to an applied knowledge test), the results could be quite different. As such, this analysis is only an illustration of the approach to modelling value-added. As it stands, there is clearly insufficient information available to explain causality for these findings.

INSERT TABLE 1 HERE

Discussion

Our results suggest that trainees in some regions may perform on average significantly better or worse than predicted, while results in other regions are not significantly different from predictions. Regarding a potential methodology for evaluating value-added, our results also suggest that a regression approach could be used to model value-added performance of training providers. However, we acknowledge that these results may raise several complex scientific, practical and political issues that, although theoretically appropriate, are challenging to resolve in high stakes settings[19,20,21]. For example, using a regression approach does not allow a researcher to infer causality and there may be several intervening variables at play that might explain observed findings within (and between) training providers. In overview, accurate evaluations of value-added can only be achieved if the value-added model and methodology is sufficiently robust and relevant for the context of postgraduate medical education and training.

In other education contexts and more generally, the use of value-added methodology has been criticised in relation to three fundamental issues. First, test scores are affected by numerous factors in addition to incoming levels of achievement. For example, in their value-add study of 87 schools in the United Kingdom, Thomas and Mortimore [20], found that only 10% of the variation in pupil examination scores was attributable to the school once background factors were taken into account (i.e. prior attainment, ethnicity, socioeconomic status, housing, mobility and level of education, etc.). Therefore, it could be argued, that the influence of a particular teacher or programme cannot be separated from the influence of those factors or other relevant (unmeasured) factors that were not controlled for in the analysis. For this reason, value-added scores should not be used to rank individual trainers or institutions. Second, test scores represent only a portion of material from a larger domain. Therefore, the credibility of a value-added model is also contingent on the validity and comprehensiveness of the assessments underlying it. Third, tests used typically measure the learning of facts and procedures (e.g such as in the applied knowledge test in this study) rather than other important goals of education, such as clinical skills in this context. In addition, attitude, engagement, and the ability to learn independently are difficult to measure using assessment or test data and, therefore, are typically beyond the scope of value-added analysis.

In terms of statistical methodology, a variety of techniques could be employed to calculate value-added measures in the context of education that are likely to vary in the sensitivity and sophistication of analysis. Historically, a key challenge for researchers has been to develop approaches that allow the statistical analysis to separate the effect of schooling or training experience on individual student outcomes and the extent to which student characteristics at the point of intake affect those outcomes. Two main approaches are typically used: standard regression, and, multilevel modelling.

The first approach was used in this study, and is a standard statistical technique for calculating the residual difference between a student or trainee's observed and expected assessment score. As mentioned previously, the observed score is a student or trainee's actual level of attainment at the end of an educational phase (such as primary or secondary) and the expected score is the level that would be predicted on the basis of his or her previous baseline attainment. Consequently the residual score is interpreted in terms of whether a student or trainee is performing above or below what would be expected on the basis of the overall statistical relationship between the baseline and outcome attainment of all students/trainees in all schools/training regions in the sample. In essence, the residual score - which ranges from a positive to negative value - provides the key statistical measure of student/trainee relative progress (i.e. the value-added). An advantage of this approach is that several factors, including baseline attainment and other student/trainee characteristics such as gender and socio-economic status (if data are available) can be used in the analysis to provide a more sensitive estimate than would result from employing baseline prior attainment measures alone. However, a disadvantage of this approach is that the unit of analysis has to be either at the level of the student (i.e. student residual scores are calculated and aggregated to school level) or the school (i.e. school residual scores are calculated). In the former case, important information about the clustering of students within a particular school is lost, and in the latter case, detailed information about individual students is lost.

The second approach, multilevel modelling, is a more sophisticated approach and is generally recognised as the most flexible tool for examining the hierarchical nature of student attainment data [22]. This technique is a generalised form of multiple regression that involves the same principle of calculating a residual value-added score. However, the clustering of students within regions, training providers and classes (a potential source of unwanted bias in the calculation of value-added scores) can also be taken into account in the analysis. Moreover, this technique also allows the unit of analysis to include both student and the training provider levels in the analysis, as well as other clustering factors such as teacher, class or region.

The multi-level modelling approach can be used to calculate unbiased and accurate estimates of training providers' value-added (residual) scores for all students, as well as the statistical significance of individual training providers' results. If data are available for consecutive student/trainee cohorts, this technique can also be employed to model trends in value-added results over time [25]. In medical education settings, calculating the cost of training interventions versus utility for postgraduate training is not straightforward, as many components are involved. Thus, in terms of further developments that build on the basic historical illustrative data provided in this paper, we propose that future research should aim to employ multilevel modeling on equivalent, more recent longitudinal medical education datasets [19].

Benefits of measuring value-added

Value-added modelling could be used to estimate the proportion of the observed variance in trainee achievement that can be attributed to an education provider, an organisation or indeed, a trainer. For example, value-added modelling could allow us to provide evidence-based answers to the following questions:

- How effective is an education provider at producing learning gains?

- What characteristics or institutional practices are associated with effective education providers?
- What are the differences in trainee achievement between education providers across regions?
- Can high performing education providers be identified so that best practice can be shared and used to inform future commissioning activities?

To our knowledge, this is the first time that empirical evidence has been assembled to provide an initial exploration of the value-added of training providers in postgraduate medical education. While it is clear that conducting analysis of value-added is theoretically and statistically complex, conceptually it is possible and it may now be timely, given the political and economic drivers for establishing standards of training and quality assurance of provision within and across regions (and relevant internationally). Value-added modelling within postgraduate medical education and training could provide benefits to: (1) inform education providers with the aim of promoting self-evaluation and learning from best practice to drive improvements across the system; (2) inform quality assurance processes, which may be tied to an education outcomes framework; (3) provide evidence to commissioners (and trainees) who hold education providers to account; (4) assist commissioners in deciding which initiatives provide the greatest return on investment for trainees and institutions; and (5) provide data on the effectiveness of training interventions and policy initiatives to identify best practice.

Caveats when considering value-added evaluations

Value-added has been extensively used in secondary education. It is important to learn the lessons from that field[23]. Value-added is a method for triggering further consideration of how and why some teaching programmes may be more effective or less effective than others. It is arguably not a means of direct comparison of effectiveness at the level of individual educators or learners, and should not be linked to performance league tables, educator performance or assessment of trainee attainment[24]. The data provided by value-added methodology are not sufficiently granular or robust to support such use. These limitations have not always been recognised by researchers and commentators of value-added methods in schools. We do not propose or support such uses of value-added methods. Value-added approaches should support efforts to enhance educational provision and interventions. They may also be used to support evaluation of innovations, alongside other evidence, but limitations of the methodology and its application should be recognised.

Importantly, postgraduate trainees in medicine provide service and so contribute to the health system whilst they train, which further complicates assessment of value-added. Another challenge is taking into account and adjusting for confounding factors. Postgraduate trainees might not all have the same educational experiences; some might take time out of training or might do some of their training in other regions. These important factors should all be taken into account. A major challenge will be in overcoming the political objections of those involved in running education and training programmes. One way of doing this might be to use value-added evaluations as confidential quality improvement tools rather than as summative judgements.

For many providers, trainees and educators, the financial and personal cost of extended training is high. Robust longitudinal evidence and analyses are a priority, so that findings can be used to inform best practice and targeted interventions that promote trainees achieving licensure or certification on time. Whilst the reasons for failure of licensure or professional examinations and extended training time are multifaceted, exploring the impact of value-added in training interventions is a potential missing link, and could directly inform our understanding of best practice in future.

Given increasing external pressures to demonstrate the effectiveness of medical education and training, there is clearly a need to understand the wider political implications of using a value-added approach. A proactive and informed dialogue within the medical education community regarding the risks, benefits and opportunities of a value-added approach should be encouraged. The data used to illustrate these examples are historical and intentionally only sample data from larger data sets is reported to illustrate the concept using a basic regression approach, without the distraction of direct comparisons at this stage. It is hoped that this will facilitate debate and further research into the concept of assessment of value-added.

References

1. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010 Dec 4;376(9756):1923-58.

2. Liu OL. Value-added assessment in higher education: A comparison of two methods. *Higher Education*, 2011; 61, 445-461.

3. Ewell PT. (2009). Assessment, accountability, and improvement: Re-visiting the tension (NILOA Occasional Paper No.1). Urbana, IL: University of Illinois and Indiana University, National Institute of Learning Outcomes Assessment.

4. Walsh K, Levin H, Jaye P & Gazzard J. Cost analyses approaches in medical education: there are no simple solutions. *Medical Education*, 2013; 1-7.

5. "Value-added" in top medical schools? MCAT/GPA as predictors of USMLE scores. http://anastomosed.wordpress.com/2013/03/13/value-added-in-top-medical-schools-mcatgpa-as-predictors- of-usmle-scores/

6. Donnon T, Oddone Paolucci E and Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine*, 2007; 82, (1), 100-106.

7. Sanders WL and Horn SP. Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education*, 1998, 12(3), 247-256.

8. Amrein-Beardsley A. Methodological concerns about the Education Value-Added

Assessment. *Educational Researcher*, 2008, 379(2), 65-75.

9. General Medical Council. *Recognising and Approving Trainers: The Implementation Plan*. 2012. London.

10. ANONYMISED FOR REVIEW.

11. Thomas SM. Assessment and the Evaluation of Institutional Effectiveness. In: Penelope Peterson, Eva Baker, Barry McGaw, (Editors), *International Encyclopedia of Education*. 2010, Vol 3, pp. 172-180. Oxford: Elsevier.

12. Patterson F, Ferguson E, Lane P, Farrell K, Martlew J & Wells A. A competency model for General Practice: Implications for selection and development. *BJGP, 2000,* 50, 188-193.

13. Irish B & Patterson F. Selecting General Practice Registrars - where next? *BJGP,* 2011, 60(580), 849-852.

14. ANONYMISED FOR REVIEW.

15. Plint S & Patterson F. The way forward for recruitment into specialty training in the UK: Identifying critical success factors. *PMJ*, 2010, 86; 323-327.

16. ANONYMISED FOR REVIEW.

17. ANONYMISED FOR REVIEW.

18. Patterson F, Lievens F, Kerrin M, Zibarras L & Carette B. (2012) Designing selection systems for medicine: The importance of balancing predictive and political validity for high stakes selection. *International Journal of Selection and Assessment*, 20(4), 486-496.

19. Organisation for Economic Co-operation and Development. (2008). *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools.* Paris: OECD.

19. Ho Kim H and Diane Lalancette D (2013). Literature Review on the value-added

measurement in higher education. *Organisation for Economic Co-operation and Development*. Paris: OECD.

20. Thomas S. Value-added measures of school effectiveness in the United Kingdom. *Prospects*, 1998, 28 (1), 91-108.

21. Raudenbush SW and Bryk AS. *Hierarchical linear models. Applications and data analysis methods,* 2002, second edition, SAGE Publications, Inc.

22. Goldstein H & Thomas S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society*, 159, 149-163.

23. Taylor J. & Nguyen AN. (2006). An analysis of the value-added by secondary schools in England: Is the value-added indicator of any value? *Oxford Bulletin of Economics and Statistics*, 68(2), 203-224.

24. Goldstein H & Spiegelhalter DJ. (1996) League tables and their limitations: Statistical issues in comparisons of institutional performance. Journal of the Royal Statistical Society, 159 (3), 385-443

25. Thomas, SM, Peng, W-J, & Gray, J. (2007) Value added trends in English secondary school performance over ten years *Oxford Review of Education*, 33 (3: 261 - 295). ISSN: 0162-3737

Table 1. Value-added using studentised residuals on the applied knowledge test (AKT) component of the MRCGP exam by UK Region in 2007

.

Training Region	Sample size for region	Studentised Residual	
		Mean	SD
16	Small	0.32*	0.96
9	Medium	0.29*	0.97
6	Medium	0.20*	1.07
14	Small	0.16	1.01
13	Small	0.07	0.96
5	Large	0.06	1.09
4	Large	0.03	1.05
3	Large	-0.04	0.93
10	Medium	-0.37*	0.98

*p<0.05, 2-tailed t-test of difference from zero. AKT = Applied Knowledge Test examination of the UK MRCGP.

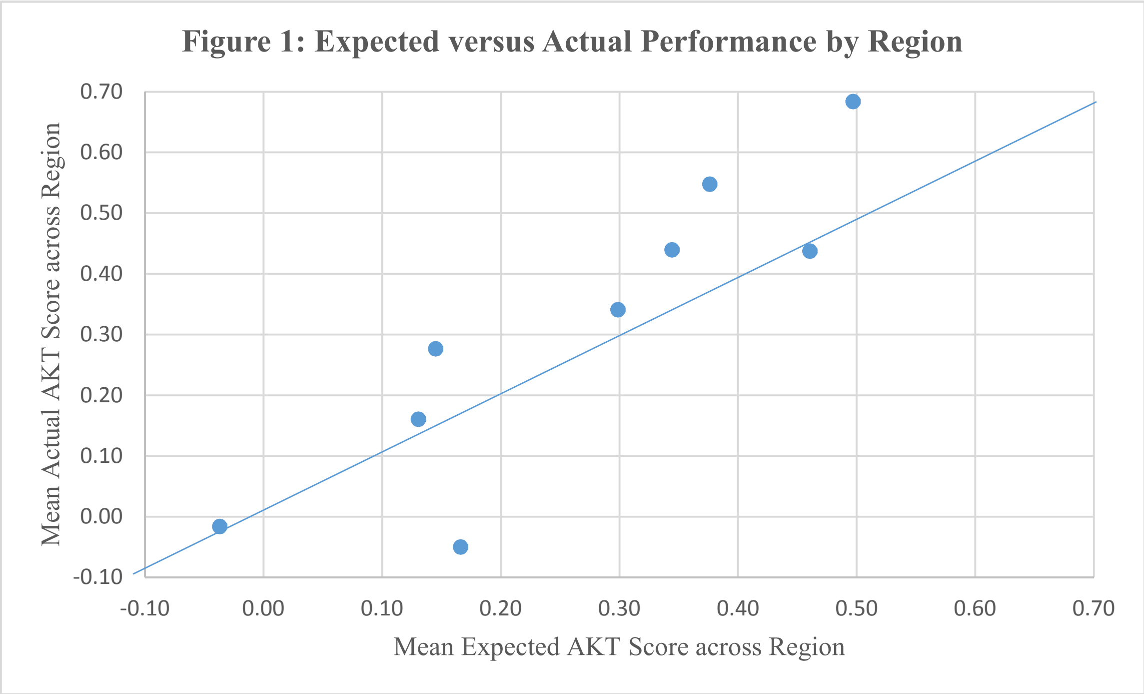
Figure 1: Expected versus Actual Performance by Region
Mean Actual AKT Score across Region
Mean Expected AKT Score across Region
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.10
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70