

# National Trainee Survey 2006 – key findings





Published by Postgraduate Medical Education and Training Board

Hercules House  
Hercules Road  
London  
SE1 7DU

Telephone: 020 7160 6100

[www.pmetb.org.uk/traineesurvey](http://www.pmetb.org.uk/traineesurvey)

ISBN: 978-0-9555910-0-6

April 2007

Report written by Daniel Smith (PMETB), Patricia Le Rolland (PMETB) and Professor  
Elisabeth Paice (London Deanery)

© PMETB 2007

# Contents

<b>Acknowledgments.....</b>	<b>6</b>
<b>Summary.....</b>	<b>7</b>
Main findings.....	7
Supervision – the key to satisfaction.....	7
Layout of this report.....	8
<b>1. Introduction.....</b>	<b>9</b>
Background.....	9
Testing the face validity of the survey.....	9
Deaneries' previous work.....	10
<b>2. Methodology and data preparation.....</b>	<b>11</b>
Data collection.....	11
Inclusion criteria.....	12
Data preparation.....	13
Response rates for the included cases.....	13
The response rate and outlying indicator scores.....	14
Derivation of the indicator scores.....	15
Method variance.....	25
Time in post.....	25
Adjustment for method variance.....	26
<b>3. Key findings.....</b>	<b>27</b>
The survey's outcome variables.....	27
Overall Satisfaction.....	28
Factors associated with job satisfaction.....	28
Medical Errors.....	31
Supervision Score.....	34
Feeling bullied – overall prevalence.....	35
The source of the perceived bullying.....	35
Reporting of bullying and subsequent action.....	36
Are certain types of trainee more likely to be bullied?.....	36
Perceived bullying and indicator scores.....	38
National training initiatives: Radiology Academies and Hospital at Night.....	39
Deaneries, training provider and differences in Overall Satisfaction and Supervision – a multilevel model analysis .....	44

<b>4. The survey data as a quality management tool.....</b>	<b>45</b>
Data releases.....	45
The identification of outliers for follow up.....	47
An example of identifying outliers – surgical specialty group.....	47
Deanery quality improvement.....	48
<b>5. The future of the National Trainee Survey.....</b>	<b>50</b>
Route of administration.....	50
The content of the survey.....	50
Adequate experience.....	51
Overall Satisfaction Score.....	51
Follow up of concerns.....	51
Personality variables.....	51
Dissemination of the results.....	52
Population data.....	53
The survey as screening tool.....	54
Making the survey mandatory.....	55
Comparison with 2006 data.....	55
National use of the data by Royal Colleges.....	55
The 2007 survey and other data sources.....	55
Deanery action plans.....	56
Indicative plan for the 2007 survey.....	56
Feedback form.....	56
<b>References.....</b>	<b>57</b>
<b>Appendices</b> – available online from <a href="http://www.pmetb.org.uk/traineesurvey">www.pmetb.org.uk/traineesurvey</a>	

# Acknowledgments

PMETB thanks the following groups of people for their help with this work:

- all the deanery survey contacts who administered the survey in their area;
- all the postgraduate medical education centre staff who assisted with the data collection;
- the deans from across the UK who supported this work.

The Surveys Working Group members:

Professor Elisabeth Paice  
Chair of COPMeD and Chair of Surveys Working Group

Maryanne Aitken  
COPMeD National Trainee Survey Manager and Quality Manager, London Deanery

Dr Gellisse Bagnall  
Educational Development Manager, NHS Education For Scotland

Dr Elaine Dolman  
Trent Deanery

Professor Shelley Heard  
Postgraduate Dean, London Deanery

Dr Stephen Kelly  
Associate Dean, West Midlands Deanery

Dr Mark Rickenbach  
Associate Dean, NHS Education South Central - Wessex Deanery

Dr Heather Payne  
Associate Dean, Wales Deanery

Surveys Working Group PMETB staff:

Mark Dexter  
Head of Policy

The contractors who assisted with data collection:

Nathan Collins                      WebLabs

Ray Flux                              CivilEyes

Bhavna Patel                        Document Capture Company

Samir Sayed                        Document Capture Company

Jane Smith                          Athene Communications

National Association of Medical Education Management

PMETB is responsible for this report's contents.

# Summary

## Main findings

Overall the majority of trainees who took part responded positively to the items they completed for the National Trainee Survey. This indicates that the majority of trainees were broadly satisfied with their training posts but there were interesting variations by provider and specialty. For example, it is possible to identify specialties and training providers that had low scores when compared to the national figures and deaneries across the UK will use this information to target areas for improvement. To ensure that the survey delivers improvements to trainees' experience of their training programmes, all UK deaneries have been using the data from the survey to prepare action plans based on their results.

PMETB's analysis has shown which facets of trainees' experience of their posts were related to the survey's outcome measures: these outcome measures were 'Overall Satisfaction' and 'Medical Errors'. The analysis shows that better supervision is associated with trainees reporting fewer perceived medical errors and greater satisfaction with their posts.

## Supervision – the key to satisfaction

The survey suggests that trainees' perception of their satisfaction and supervision vary by specialty grade group. GP trainees, for example, had the highest scores on both these indicators, suggesting that they were both better supervised and more satisfied than other trainees. Surgical group Senior House Officers (SHOs) had the lowest. Reported bullying also varied by specialty group and grade: it is interesting to note that the best supervised and most satisfied trainees – the GPs – had the lowest incidence of reported bullying.

The relationship between demographic variables and reported bullying varies by specialty. As an example, female trainees were more likely to report being bullied in anaesthetic, emergency medicine and surgical specialty groups than in others.

Survey data were also used to measure the impact of national initiatives on training. As examples, radiology trainees in academies reported higher Overall Satisfaction with their posts than those not in academies and trainees working at sites in the Hospital at Night initiative were more likely to report multidisciplinary handovers than trainees who were not.

## Layout of this report

Chapter 1 of this report covers the background to the survey. Chapter 2 covers the data collection methodology and the derivation of the indicator scores used in the screening tools that PMETB has provided to deaneries. Chapter 3 presents some of the key national findings from the data as a means to demonstrate that the indicator scores are effective measures of providers' adherence to PMETB's Generic standards for training. Both chapters 2 and 3 contain statistical tables for interested readers. Chapter 4 gives details of how the indicator scores were used in the screening tool. To avoid using the data as a league table and maintain the emphasis on quality improvement, data on individual educational providers are not given. Deaneries do have indicator score data for providers in the screening tools. Chapter 5 makes recommendations for the next trainee survey based on PMETB's evaluation of the 2006 work and feedback received from those who worked on the project.

Throughout this report indicator scores are in title case, for instance Overall Supervision. This refers specifically to the Overall Satisfaction Score derived from items on this survey, as opposed to overall satisfaction generally.

The appendices referred to throughout are available online at [www.pmetb.org.uk/traineesurvey](http://www.pmetb.org.uk/traineesurvey)

# 1. Introduction

## Background

In October 2005 PMETB inaugurated a Surveys Working Group chaired by Professor Elisabeth Paice, charged with implementing a national survey of trainee doctors. This group reports to PMETB's Statutory Training Committee. The stated purpose of the survey was to determine whether national training standards were being met. PMETB's *Generic standards for training* were developed in tandem with items for the National Trainee Survey and survey items were mapped to the standards (see Table 5, Chapter 2). However, it was not possible to measure adherence to all of the standards using a trainee survey instrument. Indeed, earlier work by Grant *et al*<sup>ii</sup> recognised that any trainee survey would need to be one part of a national data set used to quality assure postgraduate medical training. Grant *et al* proposed that this national data set could be used to inform quality assurance visits and to compare the quality of training across providers and over time.

Grant *et al* noted that Professor Paice at London Deanery had developed a *Point of view* survey to record trainees' opinions about their training<sup>iii</sup>. They recommended that it would be appropriate to use London Deanery's *Point of view* survey as the basis for the proposed trainee instrument because:

- it addressed many of the issues included in the hospital visiting documentation used by Royal Colleges and deaneries;
- its core questions were developed and validated through face-to-face interviews with over 300 SHOs;
- it had been successfully administered since 1996;
- the questions were suitable for every grade and specialty;
- it has also been used in Kent, Surrey Sussex (KSS) and Eastern deaneries.

PMETB followed this recommendation and the PMETB National Trainee Survey 2006 used many items from the *Point of view* survey.

As part of the initial development of the survey, PMETB consulted with a number of trainee groups, including the Academy of Medical Royal Colleges Trainee Doctors Group and the British Medical Association's Junior Doctors Committee. Several of the Academy group's suggestions were followed: adding an item on being pressured to submit working hours that are compliant with the European Working Time Directive (item D2 on the survey) and an item on research opportunities, and dropping an item on trainees' sexual orientation.

## Testing the face validity of the survey

PMETB commissioned researchers at the University of Winchester to recruit volunteer trainees, administer the survey to them and conduct interviews after completion of the survey to ascertain that the face validity of items was acceptable. Data were obtained from 64 volunteers working in a range of specialties from four deaneries on:

- the clarity of the items and their understanding of the items;
- whether they had any problems understanding the item;
- whether the questionnaire covered all the relevant issues.

Participants' comments were synthesised by the Winchester researchers into recommendations for the survey<sup>iv</sup>. It was possible to incorporate one of these recommendations in the 2006 survey, namely the inclusion of items on making medical errors. The items used were derived from a United States study by Baldwin and Daugherty, which looked at the effect of sleep deprivation on residents' working<sup>v</sup>. The Winchester group's remaining recommendations will be reconsidered for the next National Trainee Survey.

## **Deaneries' previous work**

The majority of deaneries (17 out of the 18 for which there are data) had already conducted surveys of trainees of one sort or another. PMETB's and COPMeD's intention continues to be that the National Trainee Survey would replace local surveys at the time of administration to avoid questionnaire fatigue among trainees and to adhere to the principles of better regulation laid out in the concordat<sup>vi</sup>. Where deanery surveys occur more frequently than annually, their surveys that do not take place when the national survey is running can continue to take place.

Differences in methods of administering the previous surveys across deaneries were apparent. Some deaneries chose to survey at a particular point in time (snapshot), while others chose to survey at the end of the trainees' posts. Owing to the fact that the PMETB and COPMeD National Trainee Survey needed to obtain data on all posts within a limited timeframe, a snapshot approach was used; this means that respondents had been in post for varying lengths of time at the point they completed the survey. The Surveys Working Group will review the merits of snapshot versus end-of-post surveys when planning future survey work. It is worth noting that the two approaches may not be mutually exclusive; for example, some trainees could be surveyed at end of post during a snapshot survey.

Queries regarding the content of the report should be sent to:

[Trainee.survey@pmetb.org.uk](mailto:Trainee.survey@pmetb.org.uk)

## 2. Methodology and data preparation

### Data collection

Data were collected between 15 May and 4 August 2006. Deaneries nominated a survey contact for their deanery to work with the PMETB/COPMeD team. Deaneries used one or more of the routes of administration detailed in Table 1 and were kindly supported by Postgraduate Medical Education Centres. Each route used the same items, with minor variation in presentation where required by the format.

**Table 1 Routes of administration** <sup>1</sup>

Route	Description	Provider	N (post-exclusions)
Portable electronic survey units (known as black boxes)	Respondents completed the survey on these units by pressing numeric keys. The units were placed in Postgraduate Medical Education Centres for a limited time-span. Postgraduate Medical Education Centre staff directed the trainees to the units.	Civil Eyes Research Limited	6,463
PMETB – web by email	PMETB set-up a website ( <a href="http://www.traineesurvey.org.uk">www.traineesurvey.org.uk</a> ). Access to the survey was controlled by an individual password that was emailed to the trainee. Reminder email was sent <sup>2</sup> .	PMETB using Forms master software	7,488
PMETB – web by letter	PMETB set-up a website ( <a href="http://www.traineesurvey.org.uk">www.traineesurvey.org.uk</a> ). Access to the survey was controlled by an individual password that was sent to the trainee by post as the deanery wished to use the website but did not have valid email addresses for the trainees.	PMETB using Forms master software	

<sup>1</sup> A summary of deaneries' data collection is available here:

[http://www.pmetb.org.uk/fileadmin/user/QA/Trainee\\_Survey/PMETBCOPMeDTraineeSurveyBriefingNote3\\_1.pdf](http://www.pmetb.org.uk/fileadmin/user/QA/Trainee_Survey/PMETBCOPMeDTraineeSurveyBriefingNote3_1.pdf)

<sup>2</sup> Some emails did not get through the first time due to respondents having set their email accounts to have enhanced junk mail filters. PMETB publicised this problem to ensure users checked their junk mail folders and resent them using an Outlook mail-merge rather than the database to get past the filter. Further consideration will be given to this issue for future surveys.

Route	Description	Provider	N (post-exclusions)
PMETB – scannable paper form	The survey was formatted to fit onto four sides of A4 (i.e. one piece of A3 paper) for ease of scanning. A separate sheet with listing specialties and their codes was provided so respondents could answer A5 and A6 using codes. Returns were scanned using Optical Mark and Optical Character Recognition.	Document Capture Company	8,634
Severn and Wessex website	A website provided by the suppliers of the Intrepid database and allowing the deanery to link response directly to their Intrepid data.	Hicomm	2,295

Data obtained through these routes were merged into one SPSS data file; this was a slow process as not all providers complied with the data template issued by PMETB. Final variables (i.e. the merged versions) were checked against the various sources to ensure that the data were not corrupted.

The items from the survey are given in Appendix 1 (available from [www.pmetb.org.uk/traineesurvey](http://www.pmetb.org.uk/traineesurvey)).

## Inclusion criteria

The focus of the survey was trainees in educationally approved posts, so the following criteria were used:

- Included: SHOs in approved posts, Specialist Registrars (SpRs), Locum Appointment Training posts (LATS), Fixed Term Training Appointments (FTTAs) and GP Registrars (GPRs).
- Excluded: Foundation Year 1 and Year 2 (F1s and F2s) (unless the deanery included these for local analysis only)<sup>3</sup>, people in non-approved posts such as clinical fellows or trust SHOs, trainees on maternity leave, trainees on out of programme experience (OOPE), non-medical public health specialists and dentists (unless dual-registered).

<sup>3</sup> Following a request from the General Medical Council to avoid confusion with their work.

## Data preparation

A total of 29,146 responses were received before the cut-off point for data entry. The following cases were excluded:

### Grade exclusions

The questionnaire included response options for grade (item A1, see online Appendix 1) for groups of trainees who were excluded in case they received the survey by mistake and for foundation trainees included on a local basis only by some deaneries. Table 2 gives the numbers of these respondents who were not included in PMETB reports.

**Table 2 Grade exclusions**

Grade	N
F1	1,669
F2	501
Trust doctor	551
Staff grade/other	739
Total	3,460

### Other exclusions

Cases were also excluded for the following reasons:

- data were missing on more than 14 items;
- there was a free text comment indicating that the respondent should not have been included, such as a reference to OOPE;
- there was a comment to say that the respondent was completing a second return.

This process left 24,880 cases available for analysis. The numbers available for individual items on the survey may be below this due to “Not applicable” response options and completion errors on the paper form.

## Response rates for the included cases

As no master population file, listing all trainees by specialty and provider, was available at the start of the survey, period data quality issues affected the calculation of response rates. These were manifest in two ways: response rates greater than 100 per cent and locations with no denominator. In addition, recording of SHO VTS (Vocational Training Scheme for GPs) was inconsistent on these returns. These problems were greater for finer aggregations, such as specialty groups, within a location. Given this, response rates are only reported by deanery here. Deaneries have received a response rate file that included the response rate for each specialty group at each acute provider; 66 per cent (214 out of 326 acute providers with population data) had overall response rates of 50 per cent or more.

## The response rate and outlying indicator scores

The provider level response file was merged with the aggregated data set and for each specialty group correlations were obtained between:

- SHO response rate and the total number of outlying indicator scores below the mean and indicator scores above the mean;
- SpR response rate and the total number of outlying indicator scores below the mean and indicator scores above the mean.

The Bonferroni correction<sup>vi</sup> was applied for multiple testing; only one of the correlations was found to be statistically significant at  $P < 0.05$ . The surgery group: the correlation between the Surgery Group SHO response rate and the total number of indicator scores lower than national mean ( $r = 0.27$ ,  $P = 0.04$ ). This suggests that the higher the response rate for this group of trainees at a provider the more negative the indicator scores for the provider, although the correlation was weak. Overall, it can be concluded that providers' response rates are not correlated with their indicator scores.

**Table 3 Response rate by deanery**

Deanery	Response rate %
Eastern	100**
Kent, Surrey, Sussex	68
Leicestershire, Northamptonshire & Rutland	64
London	64
Mersey	52
North Western	57
Northern	66
Northern Ireland	63
Oxford	63
Peninsula	65
Scotland (East)	57
Scotland (North)	97
Scotland (South East)	47
Scotland (West)	49
Severn and Wessex	74
South Yorkshire & South Humber	41
Trent	56
Wales	33
West Midlands	81
Yorkshire	65

\*\*The response rate was greater than 100 per cent; presumably either some trainees recorded their grades incorrectly or the population data were inaccurate.

## Derivation of the indicator scores

Indicator scores were derived from the items by mapping to PMETB's *Generic standards for training*. This mapping is shown in Table 5 below, together with details of how the scores were calculated. For ordinal scores, the reliability coefficients<sup>4</sup> measuring the score's internal consistency are given, which, according to Nunnally's<sup>viii</sup> thresholds, are acceptable for the following scores: Adequate Experience, Feedback, Handover and Overall Satisfaction.

Items that were ordinal in nature (that is the responses that were on scales such as "Very Poor" to "Excellent") were subjected to exploratory factor analysis<sup>5</sup> to test for construct validity<sup>ix</sup>. Items measuring a given construct should load<sup>6</sup> on the same factor and not on other factors. The results of the factor analysis are given in Table 4 below.

Using Eigen values greater than 1 as a criterion, the factor analysis supports the construction of the following indicator scores, because the items used load on the same factor and not on other factors, Overall Satisfaction, Supervision, Workload, Handover and Access to Resources.

However, other scales were more problematic; the Other Learning Opportunities and Feedback Score items did not load on their own factors and some of the items from these two scores did not load on any of the factors. The construction of both of these scores will be reviewed for the next survey.

The factor analysis failed to provide evidence of discriminant validity for the Adequate Experience Score; items D7 and D8 loaded on the same factor as the Overall Satisfaction Score items. This suggests that they are not measuring a separate construct from Overall Satisfaction, perhaps because they are also overall rating items like the H2 to H6 items that are used to derive the Overall Satisfaction Score. PMETB plans to address this issue in 2007 using specialty specific items to look at facets of experience. Rather than asking for a rating overall, the survey will ask respondents about particular facets of their placement; for instance, there may be items on theatre time for surgical trainees.

The results of the factor analysis were used to inform which items were used to derive the indicator scores but, due to the stated objective of the work, priority was given to the mapping to the *Generic standards for training* when developing the indicator scores.

<sup>4</sup> Cronbach's alpha: the mean of all possible split-half reliability coefficients

<sup>5</sup> The goal of factor analysis is to summarise the pattern of correlations among the items to reduce the number of items to a smaller number of factors.

<sup>6</sup> The loading is the correlation between the survey item and the factor, the size of the loading reflects the extent of the relationship between the item and the factor.

**Table 4 Factor analysis of ordinal survey items**

Q	Items	Overall Satisfaction	Supervision	Workload	Mixture - Learning	Handover	Access to Resources
	Variance explained	19%	8%	8%	7%	7%	6%
H4	How would you rate the quality of experience in this post?	0.88					
H5	How would you describe this post to a friend who was thinking of applying for it?	0.84					
D7	How would you rate the practical experience you are getting in this post?	0.82					
H6	How useful do you feel this post will be for your future career?	0.81					
D8	How confident are you that your current post will help you acquire the competences you need at this stage of your training?	0.78					
H3	How would you rate the quality of supervision in this post?	0.65					
H2	How would you rate the quality of teaching in this post?	0.63			0.39		
G7	In this post, how would you rate the encouragement you have had to take study leave?	0.37					
C2	How often, if ever, have you been supervised by someone who you feel isn't competent to do so?		-0.66				
C1	How often have you felt forced to cope with problems beyond your competence or experience?		-0.63				
C4	Do you always know who is providing your clinical supervision when you are working?		0.49				
C3	How often have you been expected to obtain consent for procedures which you do not carry out yourself?		-0.47				
C5	Please indicate your perception of the way in which critical events and near misses are reported in your department.		0.47				
D9	How would you rate the intensity of your work, by day?			0.76			
D10	How would you rate the intensity of your work, by night?			0.73			
D4	How often has your current working pattern left you feeling short of sleep when at work?		-0.33	0.62			
D3	How often do you work beyond your rostered hours?		-0.35	0.50			
C5*	How often do you have the opportunity to learn together with other healthcare professionals?				0.66		
C3*	Hours of relevant, timetabled, organised educational meetings or other events of educational value on average each week				0.61		
E1	How often have you had informal feedback from a senior clinician on how you are doing in this post?	0.35			0.46		
H1	How would you rate the quality of induction in this post?				0.43		
D5	Which of the following best describes handover arrangements AFTER night duty in your post?					0.89	
D6	Which of the following best describes handover arrangements BEFORE night duty in your post?					0.88	
C9	Do you have access to the Internet at your place of work?						0.76
G10	How easy is it to get access to the library services you need?						0.74

\*For presentation in this table the text of these items have been slightly edited.  
 Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalisation. Rotation converged in six iterations.  
 Only loadings greater than 0.32 are displayed. Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0.897. N = 16,609

**Table 5 The indicator scores**

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( <i>α</i> )
Access to Educational Resources	Domain 8. Educational resources and capacity	8.2 There must be access to educational facilities (including a library), and resources (including access to the Internet in all workplaces) of a standard to enable trainees to achieve the outcomes of the programme as specified in the curriculum.	G9. Do you have access to the Internet at your place of work? G10. Have you had the opportunity to participate in research in this post?	The items included here under educational resources are those that relate to individual posts. Library services relate to locations as a whole and are dealt with elsewhere.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated or if only one item is present its score is used.	0.41
Adequate Experience	Domain 5. Delivery of curriculum including assessment	5.1 Sufficient practical experience must be available within the programme to support acquisition of competence as set out in the curriculum.	D7. How would you rate the practical experience you are getting in this post? D8. How confident are you that your current post will help you acquire the competences you need at this stage of your training?	Practical experience is the <i>sine qua non</i> of postgraduate training. While D7 and D8 might in theory have elicited different responses, responses are highly correlated.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated, or if only one item is present its score is used.	0.80
Feedback	Domain 6. Support and development of trainees, trainers and local faculty	6.6 Trainees must have further meetings with their educational supervisor (or representative) at least three-monthly, to discuss their progress, outstanding learning needs and how to meet them.	E1. How often have you had informal feedback from a senior clinician on how you are doing in this post? E2. Have you had a formal meeting with your supervisor to talk about your progress in this post? E3. Have you had formal assessment of your performance in the workplace?	Feedback is an important factor in learning. This score is based on the availability of day-to-day feedback, appraisal and assessment.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated if at least two of the three items are present.	0.72

Indicator Score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( $\alpha$ )
Handover	Domain 1. Patient Safety	1.6 Trainees in hospital posts must have well organised handover arrangements ensuring continuity of patient care at the start and end of day or night duties.	D5. Which of the following best describes handover arrangements BEFORE night duty in your post? D6. Which of the following best describes handover arrangements AFTER night duty in your post?	Higher scores indicate that handover is more formally organised and more likely to be inclusive of the full multi-professional team.	Items recoded to 0 to 100 scale, where 100 is a good score, so a low score indicates a less formal handover. A mean is then calculated, or if only one item is present its score is used.	0.78
Other Learning Opportunities	Domain 6. Support and development of trainees, trainers and local faculty	6.13 Trainees must regularly be involved in the clinical audit process, including personally participating in planning, data collection and analysis. 6.17 Trainees must have the opportunity to learn with other healthcare professionals. 6.20 Trainees must be able to take study leave up to the maximum permitted in their terms and conditions. 6.25 Trainees should be exposed during their training to the academic opportunities available in their speciality.	G1. To what extent are you involved in clinical audit in this post? G2. Do you currently have access to e-learning material relevant to your training? G5. How often do you have the opportunity to learn together with other healthcare professionals (e.g. nurses, physiotherapists etc.)? G6. Have you applied for study leave in this post? G7 In this post, how would you rate the encouragement you have had to take study leave? G10 Have you had the opportunity to participate in research in this post?	This indicator combines a range of unrelated additional opportunities. A low score would indicate the need to explore which of these was problematic.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated if at least five of the six items are present.	0.36

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( <i>α</i> )
Overall Satisfaction Score	N/A internal outcome measure	N/A internal outcome measure	H2. How would you rate the quality of teaching in this post? H3. How would you rate the quality of supervision in this post? H4. How would you rate the quality of experience in this post? H5. How would you describe this post to a friend who was thinking of applying for it? H6. How useful do you feel this post will be for your future career?	This indicator combines satisfaction with each of the key elements of a training post and provides a global satisfaction score.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated if at least four of the five items are present.	0.89
Supervision	Domain 1: Patient Safety	It covers the following mandatory standards: 1.1 Trainees must make the needs of patients their first concern. 1.2 Trainees must be appropriately supervised according to their experience and competence. 1.3 Those supervising the clinical care provided by trainees must be clearly identified, competent to do so, accessible and approachable by day and by night, with time for these responsibilities clearly identified within their job plan. 1.4 Trainees must be expected to obtain consent only for procedures which they are competent to perform.	C1. How often have you felt forced to cope with problems beyond your competence or experience? C2. How often, if ever, have you been supervised by someone who you feel isn't competent to do so? C3. How often have you been expected to obtain consent for procedures which you do not carry out yourself? C4. Do you always know who is providing your clinical supervision when you are working? C5. Please indicate your perception of the way in which critical events and near misses are reported in your department.	While the quality of practical experience in a post is the factor most closely related to Overall Satisfaction, the quality of supervision is most closely related to reporting of medical errors. Good training requires good practical experience under safe supervision.	Items recoded to 0 to 100 scale, where 100 is a good score. A mean is then calculated if at least four of the five items are present.	0.54

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( $\alpha$ )
Work Intensity	Domain 6: Support and development of trainees, trainers and local faculty	6.9 Working patterns and intensity of work by day and by night must be appropriate for learning (neither too light nor too heavy).	D9. How would you rate the intensity of your work, by day? D10. How would you rate the intensity of your work, by night?	This indicator must be treated with caution: a high result is good, but a low result will not distinguish between excessive or inadequate workloads.	Items recoded as follows: 0 'Very inappropriate work intensity - too heavy or too light' 50 'Slightly inappropriate work intensity - slightly too heavy or slightly too light' 100 'work intensity about right' Then the mean of the two items is calculated, or if only one item is present its score is used.	0.47
Workload	Domain 1: Patient Safety	1.5 Shift and on-call rota patterns must be designed so as to minimise the adverse effects of sleep deprivation.	D3. How often do you work beyond your rostered hours? D4. How often has your current working pattern left you feeling short of sleep when at work? D9. How would you rate the intensity of your work, by day? D10. How would you rate the intensity of your work, by night?	Low scores are an indicator of a post where work intensity and/or long hours may lead to sleep deprivation or exhaustion.	Items recoded to 0 to 100 scale, where 100 is a good score, so a low score is a heavy workload. A mean is then calculated if at least three of the four items are present.	0.63

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha (α)
Career Advice	Domain 6. Support and development of trainees, trainers and local faculty	6.8 There must be ready access to career advice.	E4. Have you had a discussion with a senior colleague about your career plans?	Career advice is particularly important for the SHO grade. This indicator is not compared to the national mean or included in the totals below and above the national mean, because confidence intervals cannot be calculated as for some locations the standard deviation was 0.	Items recoded to 0 to 100 scale, where 100 is a good score.	n/a
Hours of Education	Domain 5. Delivery of curriculum including assessment	5.3 Trainees must be able to access and be free to attend training days, courses and other material that forms an intrinsic part of the training programme.	G3. How many hours of relevant, timetabled, organised educational meetings or other events of educational value do you take part in on average each week?	This indicator looks at the hours of weekly education. Trainees are unlikely to take into account monthly or less frequent regional training days, etc.	Mean number of hours. There is a ceiling of eight, which is labelled 8 or more.	n/a
Education Supervision	Domain 6. Support and development of trainees, trainers and local faculty	6.3 Trainees must have a designated educational supervisor. 6.4 Trainees must sign a training/learning agreement at the start of each post. 6.5 Trainees must have a logbook and/or a learning portfolio relevant to their current programme, which they discuss with their educational supervisor (or representative). 6.7 Trainees must have a means of feeding back in confidence their concerns and views about their training and education experience to an appropriate member of local faculty.	F1. Do you have a designated educational supervisor? F2. Do you have a training/learning agreement with your supervisor, setting out your respective responsibilities? F3. Are you using a learning portfolio in this post? F4. Are you using a log book in this post? F5. Have you been told whom to talk to in confidence if you have concerns, personal or educational?	This indicator is about the educational framework underpinning the post. Every element of the framework is associated with good training. Low scores suggest attention should be paid to programme management and the role of the director of medical education in ensuring structures and systems are in place.	Sum of "Yes" responses across the five items. Score can be from 0 to 5. If an item has no response, no score is calculated.	n/a

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( $\alpha$ )
Induction	Domain 6. Support and development of trainees, trainers and local faculty	<p>6.1 Every trainee starting a post or programme must attend a departmental induction to ensure they understand the curriculum, how their post fits within the programme, their duties and reporting arrangements, to ensure they are told about departmental policies and to meet key staff.</p> <p>6.2 At the start of every post within a programme, the educational supervisor (or representative) must discuss with the trainee the educational framework and support systems in the post and the respective responsibilities of trainee and trainer for learning. This discussion should include the setting of aims and objectives for the trainee to achieve in the post.</p>	<p>B1 Did someone explain your role and responsibilities in your unit or department at the start of this post? B2 Did you get all the information you needed about your workplace when you started working there? B3. Did you sit down with your supervisor and discuss your educational objectives for your current post?</p>	A good induction sets the tone for the whole post.	Sum of "Yes" responses across the five items. Score can be from 0 to 3. If an item has no response, no score is calculated.	n/a
Bullying by Consultants	Domain 6. Support and development of trainees, trainers and local faculty	<p>6.1.1 Trainees must not be subjected to, or subject others to, behaviour that undermines their professional confidence or self-esteem.</p>	<p>J1. Have you been subjected to persistent behaviour in this post that has undermined your professional confidence and self-esteem? J3. Which one of the following is the main source of this behaviour?</p>	Percentage of respondents who report being subjected to behaviour defined in J1 by consultants - their response to J3.	Bullying may also be perpetrated by other trainees, nurses, managers and even patients. This indicator only looks at consultant bullying, because this source is associated with the most stress	n/a

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( <i>α</i> )
European Working Time Directive percentage	Domain 2. Quality Assurance, Review and Evaluation	2.1 Programmes, posts, associated management, and data collection concerning trainees and local faculty must comply with the European Working Time Directive, Data Protection Act and Freedom of Information Act.	D1. Are your rostered working hours compliant with the European Working Time Directive? D2. Have you been asked to submit hours that are compliant with the European Working Time Directive, when the hours you actually work are not compliant		Percentage of respondents who answered "Yes" to D1 and "No" to D2. If an item has no response, no score is calculated.	n/a
Medical Error	N/A internal outcome measure	N/A internal outcome measure to assess the process standards relating to patient safety in domain 1 of the <i>Generic standards for training</i> .	H7 In the last month have you made a serious medical error? H8 In the last month, have you made a potentially serious medical error?	This score is difficult to interpret, as it is not measuring actual errors but whether the trainee will report having made them on this survey instrument.	Scored 1 if the respondent indicated they had made one or more than one serious and/or potentially serious medical error in the last month. Otherwise scored as 0. Respondents with missing data or who indicated they did not wish to answer were set to missing and excluded from the analysis.	n/a

Indicator score	PMETB Generic standards for training Domain	PMETB Generic standards covered	Items included	Interpretation	Calculation	Internal reliability: Cronbach's Alpha ( $\alpha$ )
Other Formal Teaching		6.16 Trainees must be able to access training in generic professional skills at all stages in their development.	G4. Have you had formal teaching since leaving medical school in any of the following?  Communication, appraisal skills, teaching skills, leadership, team working, patient safety, time management, medical ethics.	As the wording of these items did not pertain to the current post, this indicator can only be applied at deanery level and not the level of provider.	Scored 1 for each generic skill the trainee has received training in. Possible range 0 to 8.	n/a

## Method variance

### Route of administration

The data were analysed to test for any differences by route of administration (see Table 1 above). Differences did emerge. Respondents who submitted a return using the portable electronic survey units seemed more likely to give more negative responses. Some examples of this effect are given below in Table 6 for items D2 and H7. It seems likely that this effect is due to trainees perceiving the survey units as more anonymous than the web (trainees were contacted by email) and paper (bar codes were present on the paper forms), so respondents using the survey units were less inclined to give more socially desirable responses as they were confident they could not be identified.

**Table 6 Examples of method variance**

		N	$\chi^2$	p
	<b>In the last month, have you made a serious medical error? - percentage yes to once or more than once</b>			
All other routes	2.0	18,102	201.98	<0.001
Portable electronic survey units	5.5	6,365		
	<b>Have you been asked to submit hours that are compliant with the European Working Time Directive, when the hours you actually work are NOT compliant? - percentage yes</b>			
All other routes	11.9	16,563	1,332.77	<0.001
Portable electronic survey units	33.5	5,342		

### Time in post

As the survey was a snapshot, respondents had been in post for variable lengths of time. Comparing the indicator scores against the time in post item (A2), showed that trainees who had been in post longer were slightly more likely to give more positive answers.

## Adjustment for method variance

This method variance was particularly problematic because route of administration was confounded by provider and deanery. In order to make valid comparisons across providers or deaneries and be certain that differences were not due to variation in the route or the length of time in post, all the indicator scores<sup>8</sup> were adjusted as follows:

Each indicator score was regressed on to the survey route recoded (1 for portable electronic survey units and 0 for all other routes) and the length of time in post (as per A2).

An adjusted indicator score was calculated by applying the regression coefficients obtained in step 1: adjusted score = score + (survey route recoded \* B1) + (length of time in post \* B2).

There were then no differences in mean adjusted indicator scores by route or time in post.

It was not possible to adjust the indicator scores based on categorical items at the respondent level using this method. So these should be interpreted in isolation rather than comparing across providers.

<sup>8</sup> Access to Educational Resources, Adequate Experience, Feedback, Handover, Other Learning Opportunities, Overall Satisfaction, Supervision, Work Intensity and Workload

### 3. Key findings

The objective of the analysis presented in this chapter is to demonstrate that the indicator scores derived from the survey have both concurrent and construct validity<sup>x</sup> and that the scores can show differences across providers in trainees' perceptions of the quality of their training. The successful demonstration of these features of the scores shows that the survey can be appropriately used as a screening tool of quality assurance and quality management work. To demonstrate concurrent validity, analysis was undertaken using the surveys outcome variables. To demonstrate construct validity, the survey was used to assess national training initiatives. Finally, analysis was undertaken to look for differences across training providers on the indicator scores that earlier analysis had shown to be the most valid and reliable of the scores.

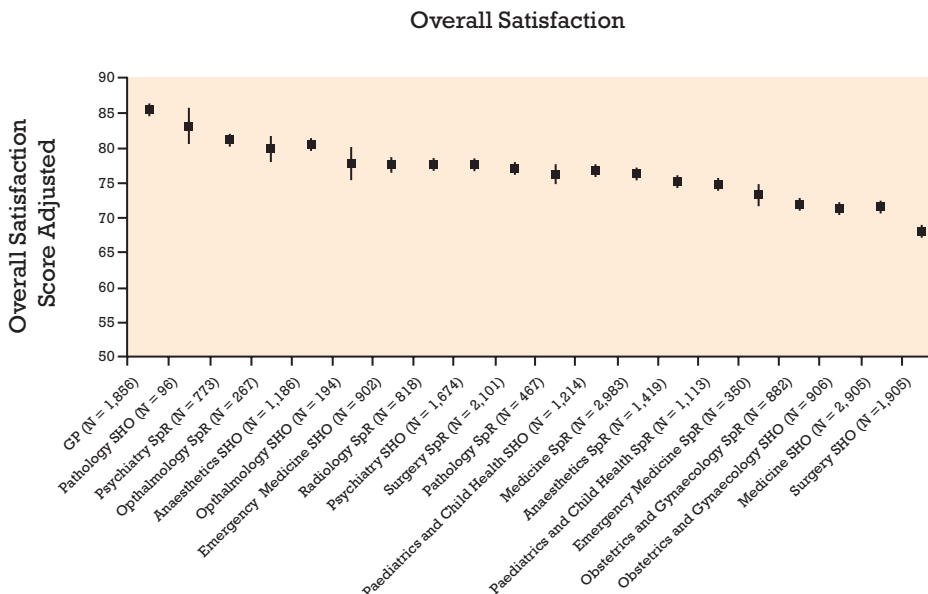
#### **The survey's outcome variables**

The survey collected a number of items that may be viewed as outcome variables: items used in the Overall Satisfaction Score and the Medical Error Score; and whether the trainee experienced behaviour that undermined their self-confidence (colloquially this might be termed bullying). Trainees' responses to these outcome variables can be predicted using the other scores from the survey to answer questions such as which aspects of trainees' perceptions of their posts are most strongly related to trainees' Overall Satisfaction Scores. If relationships between these measures of aspects of the trainees' experience and these outcomes variables are found, they provide evidence that the indicator scores have concurrent validity; i.e. trainees' perceptions of facets of their training relate to outcomes measures that were obtained at the same time as the facet measures, such as the Overall Satisfaction Score. One would expect these relationships to exist, so the object is not to demonstrate that the relationships exist (as one might do in research), but to provide evidence that the survey items are measuring what they purport to. For example, one would expect the supervision of trainees to be related to whether they report having made perceived medical errors; therefore the Supervision Score should be associated with the Medical Error Score.

## Overall Satisfaction

Trainees' Overall Satisfaction with their training posts varies by specialty and grade; this is illustrated in Chart 1<sup>9</sup>.

**Chart 1 Trainees' Overall Satisfaction Score by specialty/grade groups**



On this measure, surgical SHOs are the least satisfied with their training. Looking at the items included in the Overall Satisfaction Score, 3.9 per cent of surgical SHOs (N = 1,904) indicated that they would describe their post as very poor to a friend, compared with only 0.4 per cent of GPRs (N = 1,853). Similarly, 1.6 per cent (N = 1,907) of SHO surgeons rated their supervision as very poor compared with only 0.2 per cent (N = 1,855) of the GPs. In terms of numbers of trainees, there are 1,237 trainees in the specialties and grades on Chart 1 (5.2 per cent) who indicated that they would describe their post as poor or very poor to a friend.

## Factors associated with job satisfaction

The two scores most closely associated with trainees' Overall Satisfaction Score are Supervision and Adequate Experience; good posts are those where trainees are exposed to useful experience under good supervision. However, the Adequate Experience Score was excluded from the analysis that follows because the factor analysis (see Table 4, Chapter 2) suggested that it is not measuring a separate construct from the Overall Satisfaction Score. The two scores are highly correlated ( $r = 0.759$ ,  $N = 24,843$ ). This measurement problem may be partly due to perceptions of the adequacy of the experience gained being so inextricably bound up in trainees' overall satisfaction with post; it is not possible to be satisfied with a post that provides inadequate experience. Another plausible explanation, mentioned in Chapter 2, is that the Adequate Experience items are also overall rating type items like those used in the Overall Satisfaction Score. Plans to tackle this measurement issue are outlined in Chapter 5.

<sup>9</sup>The chart excludes Public Health and Occupational Medicine due to wide confidence intervals. GPs includes GPRs and SHO based in GP practices, but not SHO VTS trainees based in acute sites who are included in the appropriate acute specialty group.

Sequential multiple regression was used to examine which facets of the trainees' job, as measured by the survey's indicator scores, were most strongly related to their Overall Satisfaction Score. Independent variables were entered into the model in the order given in Table 7 below. At each step, the addition of the new variables significantly improved the prediction. A considerable amount of the variance (39 per cent) in the Overall Satisfaction Score is explained by the survey indicator scores entered in step three of the multiple regression. This is over and above the variance accounted for by the variables entered in steps one and two. So, for instance, one could not argue that difference by specialty group on the indicator scores are the explanation for differences in the Overall Satisfaction Score, because the model has already accounted for differences by specialty group.

**Table 7 Overall model predicting the trainees' Overall Satisfaction with their training**

Step	Type of variables	Variables	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	F change	P
1	Method variance	Time in current post Survey route	0.008	0.008	86.465	<0.001
2	Demographic variables	Grade Specialty group (dummy coded) Ethnicity <sup>10</sup> Sex Different group from intended <sup>11</sup> Forces Year qualified Where qualified	0.074	0.073	81.489	<0.001
3	Survey indicator scores	Access to Educational Resources Bullying by Consultants Educational Supervision Hours of Education Induction Other Learning Opportunities Supervision Workload	0.396	0.395	1,439.848	<0.001
N = 21,660. All specialties included						

<sup>10</sup> Ethnicity was recoded into white or non-white for the purposes of this analysis.

<sup>11</sup> Derived from A5 and A6, e.g. respondent working in medicine and intending to practise as consultant in surgery.

The following indicator scores were excluded from this model:

- Career Advice, Feedback, Handover and European Working Time scores. These were excluded because the number of not applicable/not sure/missing responses reduced the number of cases available for the analysis (all had over 3,650 respondents with missing data).
- The Work Intensity Score was removed because it contains items (D9 and D10) that were also included in the Workload Score (though coded differently) and is therefore not independent.

**Table 8 Survey variables and Overall Satisfaction**

Predictor	B Unstandardised	B Standardised	t	Sig.	sr <sup>2</sup>
Constant	97.448		2.015	0.044	
Survey route	1.283	0.038	6.838	0.000	0.001
Time in post	-0.074	-0.004	-0.632	0.527	0.000
Grade (SpR/GPR = 1)	-0.001	-0.000	-0.004	0.996	0.000
Forces (Forces =1)	-0.357	-0.002	-0.392	0.695	0.000
Anaesthetics	-1.866	-0.038	-2.962	0.003	0.000
Emergency medicine	-2.357	-0.034	-3.403	0.001	0.000
General practice	-0.208	-0.004	-0.320	0.749	0.000
Medicine	-3.194	-0.091	-5.325	0.000	0.001
Obstetrics and gynaecology	-6.329	-0.109	-9.673	0.000	0.003
Occupational medicine	-9.222	-0.034	-5.936	0.000	0.001
Paediatrics and child health	-4.687	-0.092	-7.374	0.000	0.002
Pathology	-3.402	-0.033	-4.302	0.000	0.001
Psychiatry	-3.009	-0.059	-4.752	0.000	0.001
Public health	-3.962	-0.020	-3.241	0.001	0.000
Radiology	-3.134	-0.039	-4.366	0.000	0.001
Surgery	-3.378	-0.083	-5.548	0.000	0.001
Different group from intended	-0.257	-0.006	-1.029	0.303	0.000
Year qualified	-0.035	-0.010	-1.468	0.142	0.000
Sex (female = 1)	0.571	0.019	3.325	0.001	0.000
Ethnicity (minority ethnic group = 1)	-1.551	-0.051	-7.957	0.000	0.002
Where qualified (outside UK = 1)	-1.673	-0.055	-7.762	0.000	0.002
Access to Educational Resources Score	0.051	0.055	9.697	0.000	0.003
Bullying by Consultants Score (1 = reported)	-5.817	-0.078	-14.426	0.000	0.006
Educational Supervision Score	1.149	0.095	15.583	0.000	0.007
Hours of Education Score	1.103	0.128	21.923	0.000	0.013
Induction Score	2.656	0.145	23.718	0.000	0.016
Other Learning Opportunities Score	0.177	0.209	33.856	0.000	0.032
Supervision Score	0.342	0.310	49.129	0.000	0.067
Workload Score	-0.007	-0.008	-1.321	0.187	0.000

Shared variance = 0.238, Unique variance = 0.158

Variables that are statistically significant at P < 0.05 are shaded.

Predictors that are statistically significant at  $P < 0.05$  are shaded. The direction of relationship is shown by whether a coefficient is positive or negative. All specialties are dummy coded with a 1 indicating the trainee is in that specialty and 0 indicating they are not\*.

As ethnicity is coded 1 for minority ethnic groups and the regression coefficient is negative, trainees from ethnic minorities are less likely to be satisfied with their training posts. This relationship was tested after accounting for whether the trainee qualified outside of the UK (where qualified in Table 8).

With the exception of the Workload Score the indicator scores relate to the Overall Satisfaction Score (see Table 8). This shows that as one would hope and expect they are measuring facets of the trainees' experience that relates to their satisfaction with the post. The finding that the Workload Score is not associated with the Overall Satisfaction Score, suggests that trainees do not object to working hard.

Based on the unique contribution made by the given score ( $sr^2$  in Table 8), the most important predictor<sup>12</sup> is the Supervision Score, which accounts for 6.7 per cent of the variance in the Overall Satisfaction Score after accounting for the other measures in Table 8. This finding reflects the importance of the clinical supervisors in ensuring that the trainees have a positive training experience. Trainees who perceive their supervision to be good are more likely to report being satisfied with their post. Of course, if the Adequate Experience Score is included, it is by far the most important predictor of the Overall Satisfaction Score ( $sr^2 = 0.264$ , with a  $R^2 = 0.660$ , with Supervision Score the second most important. The Adequate Experience Score is only excluded from this analysis due to measurement issues noted above.

When the scores that were excluded due to the number of cases with missing data are included in the model, together with the Adequate Experience Score (reducing the N to 10,416), the pattern remains the same: Adequate Experience is the most important followed by Supervision. The Handover Score was not related to Overall Satisfaction Score; while Career Advice, Feedback, and European Working Time Scores were all related to the Overall Satisfaction Score.

## Medical Errors

The Supervision Score is also the most important predictor of a trainee reporting that they perceived that they had made a serious or potentially serious medical error (combined into the Medical Error Score – see Table 5). Logistic regression was used to analyse which of the survey's indicator scores predicted the Medical Error Score (a binary outcome variables scored 0 or 1, see Table 5), the analysis statistically controlled for the same method and demographic variables as the multiple regression used to predict the Overall Satisfaction Score. Adding the indicator scores into the model in step 3 in Table 9 significantly improves the prediction.

<sup>12</sup> Although Adequate Experience was removed

**Table 9 Overall model predicting the trainees' reporting of medical errors**

Step	Type of variables	Variables	$\chi^2$ -step	P	$\chi^2$ -model	P
1	Method variance	Time in current post Survey route	29.33	<0.001	29.33	<0.001
2	Demographic variables	Grade Specialty group (dummy coded) Ethnicity <sup>13</sup> Sex Different group from intended <sup>14</sup> Forces Year qualified Where qualified	465.850	<0.001	495.18	<0.001
3	Survey Indicator scores	Access to Educational Resources Adequate Experience Bullying by Consultants Educational Supervision Hours of Education Induction Other Learning Opportunities Supervision Workload	334.790	<0.001	829.97	<0.001
N = 21,252. All specialties included						

**Table 10 Survey variables and reporting making a medical error**

Predictor	B	Wald	Sig.	Odds Ratio
Survey route	0.038	0.491	0.483	1.039
Time in post	-0.064	3.623	0.057	0.938
Specialty group		167.984	0.000	
Different group from intended	0.171	6.074	0.014	1.187
Grade (SpR/GPR = 1)	-0.022	0.108	0.742	0.978
Forces (Forces =1)	-0.167	0.468	0.494	0.846
Year qualified	0.022	7.310	0.007	1.022
Sex (female 1)	-0.070	1.915	0.166	0.932
Ethnicity (minority ethnic group = 1)	-0.291	25.062	0.000	0.748
Where qualified (outside UK = 1)	-0.292	18.919	0.000	0.747
Access to Educational Resources Score	-0.003	5.182	0.023	0.997
Adequate Experience Score	0.007	14.775	0.000	1.007
Bullying by Consultants Score	0.228	4.792	0.029	1.256
Educational Supervision Score	-0.001	0.001	0.981	0.999
Hours of Education Score	0.007	0.174	0.676	1.007
Induction Score	-0.063	3.998	0.046	0.938
Other Learning Opportunities Score	0.002	1.250	0.264	1.002
Supervision Score	-0.022	125.957	0.000	0.978
Workload Score	-0.012	67.402	0.000	0.988
Constant	-43.991	7.239	0.007	0.000

Variables that are statistically significant at P < 0.05 are shaded.

<sup>13</sup> Ethnicity was recoded into white or non-white for the purposes of this analysis.

<sup>14</sup> Derived from A5 and A6, e.g. respondent working in medicine and intending to practise as consultant in surgery.

In interpreting these findings, it should be noted that the variable being predicted is the trainees' reporting of perceived medical errors in the last month on this survey instrument (and not necessarily their reporting of any incident through the appropriate local procedures for reporting incidents), i.e. not actual medical errors. It is likely that the actual rate of perceived medical errors is higher than reported on the survey forms, given the method variance finding (see Table 6, Chapter 2).

Predictors that are statistically significant at  $P < 0.05$  are shaded on Table 10. The direction of relationship is shown by whether a coefficient is positive or negative. There are a number of findings of note. Trainees from a minority ethnic group and trainees who qualified outside the UK are less likely to indicate they made a medical error in the last month. Trainees who report being bullied are more likely to report making an error.

It is likely that the effect for specialty group is related to the type of work. For instance, as one might expect, emergency medicine trainees report having made more errors (16.8 per cent,  $N = 1,217$ ) than trainees in other specialties (9.3 per cent,  $N = 22,829$ ).

Higher Induction, Supervision and Workload Scores are associated with trainees being less likely to report making medical errors. This means that a clearer and more complete induction, better clinical supervision and not being overloaded with work<sup>15</sup> are associated with trainees being less likely to report making medical errors. A higher Adequate Experience Score is associated with trainees more likely to report errors: a job offering plenty of experience is related to trainees being likely to report making errors.

Given the importance of the Supervision Score and the Workload Score in predicting the reporting of medical errors, the individual items were examined; some of these are presented below in Tables 11 and 12. Trainees who report making medical errors are more likely to report saying they are forced to cope with problems beyond their competence and that they feel that they were supervised by someone not competent to do so. They are also more likely to report working beyond their rostered hours and feeling sleep deprived.

**Table 11 Medical errors and Supervision Score items**

How often have you felt forced to cope with problems beyond your competence or experience?	Never %	Rarely %	Monthly %	Weekly %	Daily %	N
No medical errors reported	21.4	60.2	10.1	7.1	1.3	21,937
One or more serious or potentially serious medical errors reported	11.2	55.2	16.2	14.2	3.1	2,339
How often, if ever, have you been supervised by someone who you feel isn't competent to do so?						
No medical errors reported	56.4	36.6	3.9	2.4	0.8	21,903
One or more serious or potentially serious medical errors reported	43.9	43.6	6.8	4.7	1.0	2,338

<sup>15</sup> All Scores were calculated so that a high score was good

**Table 12 Medical Errors and Workload items**

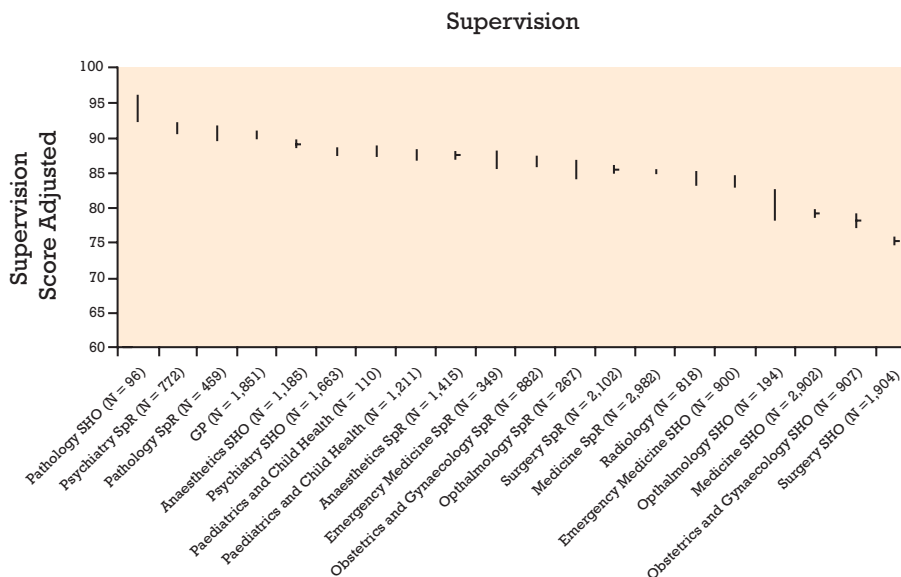
How often do you work beyond your rostered hours?	Never %	Rarely %	Monthly %	Weekly %	Daily %	N
No medical errors reported	8.4	33.8	12.7	31.6	13.5	21,904
One or more serious or potentially serious medical errors reported	5.1	22.2	13.4	37.9	21.3	2,338
How often, if ever, have you been supervised by someone who you feel isn't competent to do so?						
No medical errors reported	21.6	41.1	18.9	15.0	3.4	21,839
One or more serious or potentially serious medical errors reported	12.0	33.6	23.7	23.6	7.1	2,332

PMETB in conjunction with members of the Survey Working Group will publish a more detailed paper analysing these data.

### Supervision Score

The Supervision Score has been shown out of all the indicator scores to be the most strongly related to Overall Satisfaction Score, Medical Errors Score and Consultant Bullying Score. Like the Overall Satisfaction Score, it varies by grade and specialty.

**Chart 2 Supervision Score by specialty/grade groups**



To understand what the differences in the Supervision Score mean, one can look at some of the individual items included in its derivation: 9.7 per cent (184/1,902) of surgical SHOs, who as a group are the least satisfied with their supervision, indicated that they were supervised by someone they felt was not competent to do so, monthly or more frequently, whereas only 3.0 per cent (23/772) of the psychiatry SpRs felt that this was the case. Of the surgical SHOs, 26.6 per cent (507/1,909) reported feeling forced to cope with problems beyond their competence or experience monthly or more frequently, compared with 8.4 per cent (65/775) of the psychiatry SpRs.

## Feeling bullied – overall prevalence

To ascertain whether trainees felt bullied, the questionnaire included an item previously used<sup>xiii</sup>, “J1. Have you been subjected to persistent behaviour in this post that undermined your professional confidence and/or self-esteem?”

Overall, 10.5 per cent of trainees (N = 23,198) reported being subjected to persistent behaviour in their current post that undermined their professional confidence and/or self-esteem. A further 1,460 trainees did not wish to answer and another 222 left the item blank<sup>16</sup>, suggesting that the rate may be as high as 16.5 per cent, were these responses taken as yes.

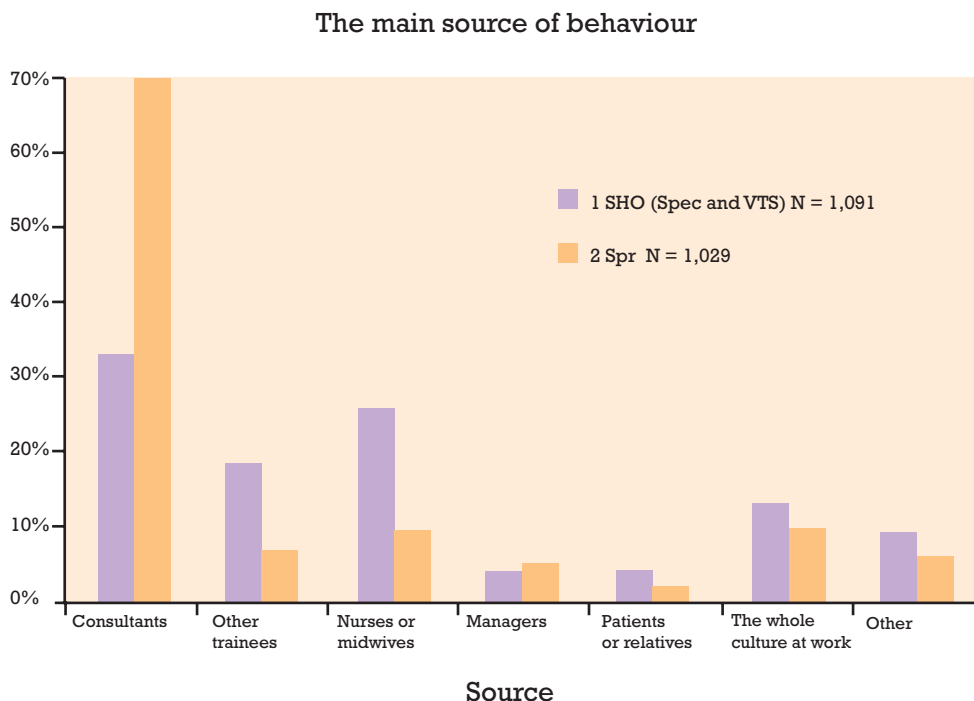
The rate is lower than that reported by Paice *et al*<sup>xiv</sup> who also used the Hicks item. On their survey, 18.1 per cent of hospital based trainees reported experiencing bullying in their current post, compared with 10.9 per cent of hospital based trainees on this survey (N = 20,947, GPRs excluded).

## The source of the perceived bullying

Respondents who indicated they had been subjected to this type of behaviour were asked to indicate its main source. Among hospital trainees (including VTS trainees on hospital placements) this varied by grade. SpR trainees were more likely to report that a consultant was the source, while SHO grade trainees were more likely to report that other trainees or nursing staff were the source (see Chart 3):

$\chi^2 = 307.4, P < 0.001$ .

**Chart 3 The source of the perceived bullying**



<sup>16</sup> This was possible on the paper route.

## Reporting of bullying and subsequent action

Of the trainees who reported bullying, 35.1 per cent (N = 2,268 – missing data in 167 cases) indicated that they reported it to their employer/someone in authority; of these, 41.7 per cent (N = 575 – missing data in 221 cases) indicated that successful action was taken to stop it. Of all the reported incidents (all incidents including those for which there are missing responses to the reporting and successful action items) only 9.9 per cent (N = 2,435) were reported and subjected to remedial action.

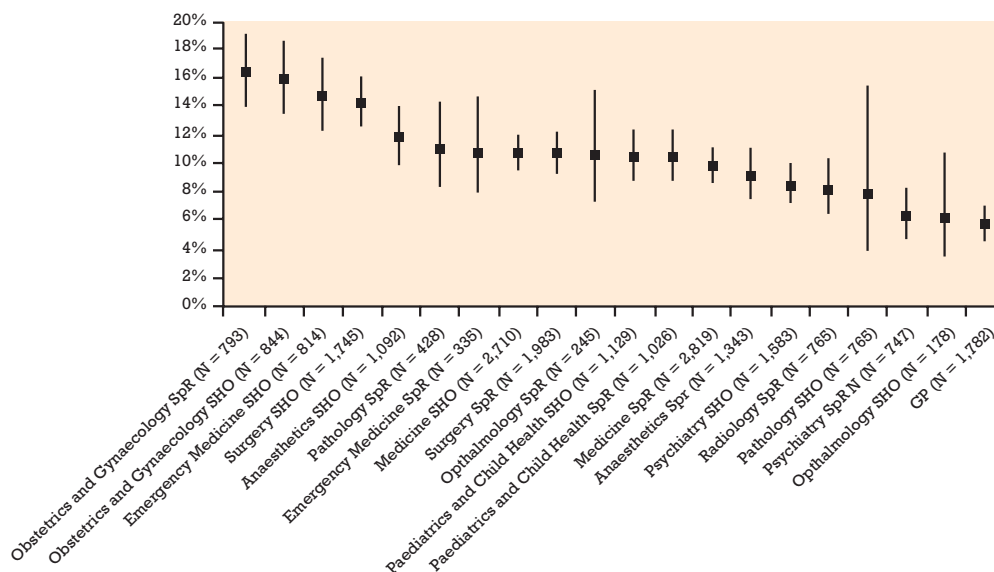
Whether a trainee who experienced bullying reported it varied according to the source of the bullying. Trainees were more likely to report it if it came from managers or patients/relatives than if it came from other clinical staff ( $\chi^2 = 21.41$ ,  $P < 0.001$ ,  $N = 1,757$  cases of bullying with a specific source only). 34.9 per cent of bullying involving trainees, consultant or nursing staff was reported, compared with 54.0 per cent of bullying involving managers or patients/relatives.

## Are certain types of trainee more likely to be bullied?

Trainees' reporting of bullying also varies by specialty and grade, as illustrated in Chart 4 below ( $\chi^2 = 175.37$ ,  $P < 0.001$ ,  $N = 22,450$ )<sup>17</sup>. The obstetrics and gynaecology trainees report the most bullying, while GPs report the least.

### Chart 4 Bullying by specialty/grade group

Have you been subjected to persistent behaviour in this post that has undermined your professional confidence and/or self-esteem?



Confidence interval calculated following Altman *et al*<sup>xv</sup> (2000)

These findings concur with existing literature that has shown an increased prevalence of “harassing or discriminatory behaviours” in surgery and obstetrics and gynaecology<sup>xvi</sup>.

Logistic regression was used to look at the relationships between feeling bullied and demographic variables. These relationships varied by specialty group (see Table 13).

<sup>17</sup> For clarity on the chart, very small or non-existent groups were excluded: public health, occupational medicine, and radiology SHOs.

**Table 13 The relationship between demographic variables and reporting bullying for each specialty group and overall**

	Variables in model. If the variable is related (P <0.05) to reporting bullying the Odds ratio is reported, if it is not related the cell is blank.							
	N	Time in post	Trainee is locum	Trainee in HM forces	Year qualified	Qualified outside UK	Female	Trainee not white
Anaesthetics	2,284						1.51	
Emergency medicine (model not significant at P <0.05)	1,099						1.56	
General Practice	1,754			4.11				1.82
Medicine	5,287	1.23						1.34
Obstetrics and gynaecology	1,535		1.99		1.04			1.71
Ophthalmology	401		11.59					
Paediatrics and child health	2,063	1.41						
Pathology (model not significant at P <0.05)	477							
Psychiatry (model not significant at P <0.05)	2,176					1.59		
Radiology	744	1.98						1.92
Surgery	3,489	1.20				1.33	1.56	
All specialties predicting any bullying.	21,758	1.21			1.01	1.15	1.11	1.30
All specialties - predicting consultant bullying	21,758	1.50	1.42		0.95	0.74		

Table 13 gives the odds ratios for each variable that is related to trainee reporting bullying at P < 0.05. Where no odds ratio is given, the relationship is not statistically significant. For instance, for surgical trainees the following variables are associated with an increased probability of reporting bullying on this survey: being in post longer, qualifying outside the UK and being female. None of the other variables are associated with reporting bullying.

With the exception of time in post and year qualified, all variables are binary, where they are coded 1 for the presence of the description given in Table 13 (e.g. female = 1). The Odds ratio for time in post and year qualified indicate that the longer the trainee has been in post, the more likely they are to report being bullied in that post and trainees who qualified more recently (i.e. left medical school recently) are more likely to report being bullied.

For binary variables, the odds ratios show the increased probability of reporting bullying if the factor is present. For all specialties, trainees from a minority ethnic group are 1.3 times more likely to report being bullied than white trainees.

In logistic regression, when all the variables are entered at the same stage, each variable is tested after accounting for all the other variables in the model, so qualifying outside the UK and being a non-white trainee increases the probability of being bullied. This is illustrated in Table 14.

**Table 14 Trainees' ethnicity and place of qualification and bullying**

	% reporting bullying	N
White qualifying within the UK	8.6	9,966
White qualifying outside the UK	9.6	1,569
Minority ethnic group qualifying within the UK	10.8	3,409
Minority ethnic group qualifying outside the UK	11.9	7,195

Looking at reports of feeling bullied by consultants only, in the final row of Table 13, gives a different pattern; neither the trainees' sex nor their ethnicity (white versus non-white) is associated with the trainees reporting bullying by a consultant. Furthermore, trainees from outside the UK are less likely to feel bullied by consultants (odds ratio = 0.74) and older trainees are more likely to report feeling bullied by consultants (this fits with the finding that SpR are more likely to feel bullied by consultants – see Chart 3).

This pattern of results reflects trainees' own perceptions. Of those who reported being bullied and answered whether this was related to their ethnicity, sex, sexual orientation or religious beliefs, 36.1 per cent of trainees who felt bullied by any non-consultant source felt it was related to one of these, against 28.3 per cent of those who reported the source as being a consultant (N = 1,888,  $\chi^2 = 12.8$ ,  $P < 0.001$ ).

The model varies across specialties; the trainee's gender is only related to perceived bullying within certain specialty groups - anaesthetics, emergency medicine and surgery. GP trainees working in the armed forces are more likely to report perceived bullying than civilian GPs (seven of the 22 GPs in the armed forces answered yes to item J1).

### Perceived bullying and indicator scores

Respondents who reported being bullied had lower scores on all the indicator scores, except the Handover Score (all mean comparisons were statistically significant at P 0.05 after the Bonferroni correction was applied), suggesting that experiencing bullying is associated with a generally less positive training post. The largest association was with the Supervision Score, again indicating the importance of this indicator.

PMETB in conjunction with members of the Survey Working Group will publish a more detailed paper analysing these data. It is worth noting here that, as Paice *et al* note, some of the behaviours that erode trainees' professional confidence or self-esteem may be attempts to improve trainees' performance, so an educational rather than punitive approach is required to tackle the problem.

## National training initiatives

The survey data were used to assess the impact of two national initiatives that have sought to improve the training experience of junior doctors:

- Radiology academies<sup>xvii</sup>
- Hospital at Night<sup>xviii</sup>

In both cases, some of the predicted differences emerged, providing further evidence of the construct validity of the scores derived from the survey data and highlighting their potential for evaluating other initiatives that are designed to improve junior doctors' training posts.

### Radiology academies

Three academies were delivered as part of the 'Radiology - Integrated Training Initiative (R-ITI)', a national programme to provide an increased number of high quality radiologists.

**Table 15 Respondents by radiology academy**

Academy	Hospitals	N
Leeds and West Yorkshire Radiology Academy	Leeds Teaching Hospitals NHS Trust	24
Norfolk & Norwich Radiology Academy	Norfolk & Norwich University Hospital NHS Trust	16
Peninsula Radiology Academy	Plymouth Hospitals NHS Trust	22
	Royal Devon & Exeter NHS Foundation Trust	1
	South Devon Healthcare NHS Trust	5

### Numbers of respondents

Only trainees in clinical radiology (also known as diagnostic radiology and formerly known as radiology) were included. The four clinical radiology trainees whose grade was recorded as something other than SpR were excluded. This left 68 trainees in the academies (as detailed in Table 15) and 536 trainees not in the academies. Trainees at the academy sites but not in the 2005 or 2006 cohorts of trainees have also been classified as training at an academy.

### Approach to analysis and results

The mean indicator scores (derived from survey items – see Table 5 in Chapter 2) for the two groups were compared and the ANOVA (F-test) used to test for differences; the Bonferroni correction for multiple testing was applied. A number of the mean scores were significantly different at  $P < 0.05$ ; they are shaded in Table 16 below. All the differences are in the expected direction. Trainees based within an academy reported a more positive training experience. The survey was not specifically designed to evaluate the radiology academies and it is likely that a survey with items concerned specifically with the training of radiologists would have found more marked differences.

**Table 16 Comparison between academy and non-academy trainees on the National Trainee Survey Indicators**

	Mean: non Academy	N: non Academy	Stan Dev: non Academy	Mean: Academy	N: Academy	Stan Dev: Academy	F	P (with Bonferroni)
Access to Educational Resources Score	79.95	536	14.18	83.02	68	12.39	2.90	1.00
Adequate Experience Score	76.91	536	13.01	74.90	68	11.75	1.47	1.00
Career Advice Score	81.76	403	36.43	85.34	58	33.78	0.50	1.00
Bullying % reporting	4%	536	20%	3%	68	17%	0.28	1.00
Education Supervision Score	3.99	535	1.15	3.96	68	1.11	0.04	1.00
European Working Time Directive % "Yes"	83%	444	38%	84%	56	37%	0.02	1.00
Feedback Score (Adjusted)	50.90	436	23.76	53.78	57	20.61	0.76	1.00
Handover Score (Adjusted)	28.86	276	18.41	33.01	23	14.93	1.11	1.00
Hour Education Score	3.34	535	1.87	4.75	68	2.70	30.66	0.00
Induction Score	2.56	534	0.76	2.51	68	0.80	0.21	1.00
Other Learning Opportunities Score (Adjusted)	51.43	521	15.21	51.66	67	16.39	0.01	1.00
Overall Satisfaction Score (Adjusted)	77.52	536	12.67	82.35	68	9.90	9.14	0.04
Supervision Score (Adjusted)	83.58	536	13.58	88.99	68	12.20	9.76	0.03
Work Intensity Score (Adjusted)	89.64	535	19.46	88.78	68	22.09	0.11	1.00
Workload Score (Adjusted) mean score at location	58.47	534	14.44	64.78	68	17.52	10.92	0.02

Significant differences in mean scores,  $P < 0.05$  are shaded above.

## The Hospital at Night

The Hospital at Night initiative aims to reduce dependency on training grade doctors for providing cover at night, in order to reduce their working hours and ensure that these are compliant with the European Working Time Directive<sup>18</sup>, while ensuring that there is no negative impact on their training. Hospital at Night advocates supervised multi-disciplinary handover in the evenings<sup>xix</sup>.

The impact of the Hospital at Night programme was assessed by comparing trainees working in hospitals with Hospital at Night teams with trainees in other acute hospitals that do not have these teams. Locations were classified as participating in Hospital at Night on the basis of data supplied by the Hospital at Night team<sup>19</sup>. On the basis of Hospital at Night's stated objectives, particular indicator scores and items were compared across these two groups of trainees.

**Table 17 Handover**

Hospital at Night respondents should report having multi-disciplinary handovers. These data suggest that this is the case (Table 17 below): 23.7 per cent reported that nurses are involved in handovers at Hospital at Night locations compared with 11.7 per cent at non-Hospital at Night providers. The difference is more marked if the comparison is restricted to medical trainees.

	Implementation status	Which of the following best describes handover arrangements BEFORE night duty in your post?						$\chi^2$	P (Bonferroni applied)
		None %	Informal %	A phone or email communication %	An organised meeting of doctors %	An organised meeting of doctors and nurses %			
All trainees at acute sites	Not implemented N = 11,388	1.9	38.5	4.2	43.7	11.7	378.63	0.00	
	Implemented N = 4,028	1.9	28.0	3.8	42.7	23.6			
Medical trainees at acute sites	Not implemented N = 3,393	1.2	44.4	5.2	32.0	17.1	507.07	0.00	
	Implemented N = 1,228	2.0	20.0	5.5	24.3	48.2			
Surgical trainees at acute sites	Not implemented N = 2,266	1.6	35.5	5.3	49.9	7.8	19.33	0.00	
	Implemented N = 725	2.2	30.1	5.5	49.8	12.4			

<sup>18</sup> <http://www.dh.gov.uk/en/Policyandguidance/Humanresourcesandtraining/Workingdifferently/Europeanworkingtimedirective/index.htm>.

<sup>19</sup> Data kindly supplied by Gerry Bolger, Project Director - National Hospital at Night Team

The difference between night to day handover between trainees at Hospital at Night sites compared with trainees at other acute sites was far less marked. The biggest difference was for the medical trainees; for this group Hospital at Night sites had 2.7 per cent more respondents saying their handover involved nurses. As the focus of Hospital at Night is on the day to night handover, these data would seem to reflect changes to handover process related to Hospital at Night implementation.

**Table 18 Impact on working times**

The Hospital at Night programme aims to help trusts comply with the European Working Time Directive. The data below do not suggest this, as the difference in compliance is in the opposite direction to that predicted and the differences, although small, are statistically significant for all trainees at acute sites and the sub-group of medical trainees at acute sites. Similarly, there are small statistically significant differences on the second item for all trainees and the sub-group of medical trainees; trainees working at Hospital at Night sites are slightly more likely to have been asked to submit hours that are compliant, when the hours actually worked were not compliant..

	Implementation status	Are your rostered working hours compliant with the European Working Time Directive?				Have you been asked to submit hours that are compliant with the European Working Time Directive, when the hours you actually work are NOT compliant?			
		% Yes	N	$\chi^2$	P (Bonferroni)	% Yes	N	$\chi^2$	P (Bonferroni)
All trainees at acute sites	Not implemented	89.0	10,834	7.160	0.045	18.9	11,716	16.064	0.000
	Implemented	87.4	3,750			21.8	4,010		
Medical trainees at acute sites	Not implemented	88.4	3,086	9.145	0.015	19.9	3,371	7.080	0.047
	Implemented	84.9	1,132			23.6	1,210		
Surgical trainees at acute sites	Not implemented	82.0	2,150	0.090	1.000	24.5	2,371	0.023	1.000
	Implemented	81.4	668			24.8	725		

Obviously this finding is only an association. It may be the case that Hospital at Night providers knew there was a problem in complying with the European Working Time Directive and, as a result, have started to tackle it, but the changes have not yet bedded down sufficiently to impact upon working times. The analysis can be repeated with the 2007 data to test this hypothesis and would be strengthened by more detailed data on the implementations of the initiative at Hospital at Night providers

**Table 19 Work intensity**

The work intensity items reflect a similar pattern, with trainees at Hospital at Night sites reporting a heavier workload during the night. Again, the differences are small but, due to the sample sizes available, they are statistically significant. The differences in workload during the day are less reliable; the difference is only statistically significant when all trainees are included in the analysis.

		How would you rate the intensity of your work by:							
		Day				Night			
		% responding heavy or very heavy	N	$\chi^2$	P (Bonferroni applied)	% responding heavy or very heavy	N	$\chi^2$	P (Bonferroni applied)
All trainees at acute sites	Not implemented	37.6	13,290	28.08	0.00	39.1	12,132	49.05	0.00
	Implemented	41.9	4,651			44.8	4,163		
Medical trainees at acute sites	Not implemented	41.0	3,829	7.99	0.55	47.0	3,460	15.65	0.02
	Implemented	44.8	1,380			52.7	1,219		
Surgical trainees at acute sites	Not implemented	34.4	2,700	2.70	1.00	25.1	2,597	15.37	0.02
	Implemented	34.1	853			31.2	804		

Chi-squared test run across all five response options. Presented as % heavy/very heavy in this table for ease interpretation only

**Table 20 Impact on training**

The Hospital at Night initiative is not meant to impact on doctors' training and the survey data suggest that, on balance, it does not affect trainees' perceptions of the experience they are getting, as there are no consistent statistically significant differences between trainees at sites that have or have not implemented Hospital at Night on the Adequate Experience Score (derivation of scores is outlined in Table 5, Chapter 2).

		Adequate Experience Score				
		Mean	SD	N	F	P (Bonferroni applied)
All trainees at acute sites	Not implemented	73.54	16.06	13,330	4.02	0.14
	Implemented	72.99	16.63	4,657		
Medical trainees at acute sites	Not implemented	72.58	15.32	3,842	0.09	1
	Implemented	72.73	16.27	1,384		
Surgical trainees at acute sites	Not implemented	70.35	19.30	2,702	5.95	0.04
	Implemented	68.49	19.90	853		

## **Deaneries, training provider and differences in Overall Satisfaction and Supervision – a multilevel model analysis**

These data are inherently multilevel: trainees work within departments, which are within hospitals that fall within a given deanery's responsibility. PMETB consequently commissioned multilevel modelling<sup>x</sup> to analyse these data at different levels<sup>xi</sup>. This analysis looked for differences between deaneries, training providers and specialty groups within training providers on two of the indicator scores: Overall Satisfaction and Supervision.

The model included a number of background variables to statistically control for their effects, so that differences between providers cannot be attributed to differences between them on these background variables.

The model has four nested hierarchical levels that reflect the organisation of postgraduate medical education: trainees; within specialty groups within providers (e.g. Medicine at a given provider); within providers; within deaneries. Significant variation was found on the Overall Satisfaction Score and the Supervision Score at the level of specialties within providers ( $P < 0.001$  and  $P < 0.001$ ) and providers ( $P = 0.001$  and  $P = 0.005$ ). There is no difference on either measure at the level of deaneries. The differences do not reflect differences in trainee mix in terms of the background variables: specialty, sex, grade, type in post, years qualified, or route of responding to the questionnaire.

While these results show that there are differences in trainees' perceptions of the quality of their training associated with training providers and specialties within training providers, it is not possible to use these data alone to reliably identify poorly performing training providers. This suggests that, while the survey data may have utility as a screening tool as proposed by PMETB this year, it would be inappropriate to take action or identify poorly performing training providers on the basis of these data alone.

## 4. The survey data as a quality management tool

As the primary objective of the survey work has been quality improvement, PMETB has focused on releasing data to deaneries, which are responsible for local quality management. Deaneries have been working with training providers within their area on action plans in response to the survey's findings. Table 21 details the data released to deaneries that was used for this work.

### Data releases

**Table 21 Data that PMETB has released**

Data	Release data and format	Audience
<p><b>Data by specialty group and provider</b></p> <p>PMETB/COPMeD National Screening Tool with indicator scores for each specialty group for all providers in the UK with more than three respondents derived from the survey data.</p>	<p>5 December 2006</p> <p><i>Compare CD</i> –</p> <p><i>Compare</i> software allows the user to explore these data and produce reports as Word documents for a given location and specialty group. The user can choose which other locations/VTSs to compare their chosen location/VTS with on the charts.</p>	<p>Deaneries</p>
<p><b>Specialty group data by provider benchmarked to national means and quartiles</b></p> <p>The indicator scores (used on the <i>Compare CD</i> above) with comparison with the national means and quartiles for the given specialty group; these means and quartiles use the data from all respondents, including those where there are less than three respondents.</p>	<p>7 December 2006</p> <p>Screening Tool Access Database that allows the user to obtain a report for a group of indicators where scores for all available specialty groups are displayed for each location.</p>	<p>Deaneries</p>
<p><b>Specialty data</b></p> <p>Indicator scores aggregated to specialty and grade groups by deanery.</p> <p>This dataset contains indicator scores for each specialty and grade group in each deanery, where there are more than three respondents available, e.g. scores for cardiology SpRs in London Deanery. This report will therefore include specialties such as public health and occupational medicine that, due to low numbers at individual providers, have not been included in data by specialty group and location.</p>	<p>11 January 2007</p> <p><i>Compare CD</i></p>	<p>Royal Colleges</p>

Data	Release data and format	Audience
<p><b>Specialty data – UK wide</b></p> <p>Indicator scores aggregated to specialty and grade for the whole of the UK.</p> <p>This dataset contains indicator scores for each specialty and grade group across the UK, where there are more than three respondents available, e.g. scores for cardiology SpRs UK wide.</p>	<p>11 January 2007</p> <p>Compare CD</p>	<p>Royal Colleges</p>
<p><b>Provider profile report</b></p> <p>A summary of each provider across all their specialty groups. It identifies the providers with the highest proportion of outlier indicator scores (both above and below the mean) across all the specialty groups.</p>	<p>1 February</p> <p>Excel</p>	<p>Deaneries</p>
<p><b>Deanery level analysis of training since leaving medical school</b></p> <p>An analysis based on response to item G4 by deanery.</p>	<p>21 December 2006</p> <p>Excel</p>	<p>Deaneries</p> <p>NHS Institute for Innovation and Improvement</p>
<p><b>Intended specialty analysis</b></p>	<p>18 December 2006</p> <p>Excel spreadsheet available here:  <a href="http://www.pmetb.org.uk/index.php?id=intendedspecialtyanalysis">http://www.pmetb.org.uk/index.php?id=intendedspecialtyanalysis</a></p>	<p>Trainees</p>

Deaneries have been using the data released to them for their own quality management activities. PMETB has not been prescriptive about how deaneries should go about this and generally deaneries have discussed and disseminated the data locally as they deem appropriate. It is important for deaneries to undertake this activity, as a national organisation such as PMETB would not be able to interpret the data within the context of the NHS locally. When considering the reason for a low score on an Indicator Score derived from survey data, it is advisable to exclude causes such as particular local circumstances (for instance a change to the configuration of services that has been disruptive) before suggesting the score is indicative of a performance issue. Only deaneries would be able to take such causes into consideration when planning follow up on the survey data.

## The identification of outliers for follow up

PMETB only intends the survey data to be used as an initial screening tool, to identify areas that may require further investigation. For the purposes of identifying outliers, the following analysis was conducted.

The survey data were aggregated to the level of specialty groups within providers, so for example there were mean adjusted<sup>20</sup> scale scores for surgical trainees at a given provider. To ensure any given trainee's responses remained anonymous as promised, scores were only calculated where there were more than three respondents working in the specialty group at the given provider. This is data release, data by specialty group and provider, in Table 21.

The mean scores for each provider's specialty group of trainees were compared with the national mean and quartiles for that specialty group (as per specialty group data by provider benchmarked to national means and quartiles in Table 21). The national comparison group included all trainees within that specialty group, including those working at providers with less than three respondents. So, for instance, the mean score for the given provider's surgical trainees was compared with the national mean and quartiles for all surgical trainee respondents across the UK. A score was defined as outlying if it was in the bottom quartile and below the national mean (based on the confidence intervals not overlapping). The advantage of including data from the confidence intervals in the comparison is that, for a sample from a particular provider, the confidences are narrower if there are more respondents *and* if there is more agreement among those who have responded. Just using the quartile information takes no account of sampling error and just using a comparison of the means meant that it was possible for providers not in the lowest quartile to be deemed outliers because they had very narrow confidence intervals. Therefore using both pieces of information was felt to be more appropriate.

## An example of identifying outliers – surgical specialty group

For the all the surgical group trainees in the UK (N = 4,022), the mean Supervision Score is 80.58 (95 per cent confidence intervals 80.11 to 81.06). The 25th percentile is 71.09. The 209 providers for whom there were four or more respondents were then compared with these figures. Table 22 shows which providers were classified as outliers because they met both criteria, and the providers that only met one of the two criteria and were therefore not classified as outliers. Providers 1 to 4 are outliers; providers 5 to 13 are not.

**Table 22 Surgical providers in the bottom quartile and/or below the national mean based on the confidence intervals**

Deanery	Provider	Mean	N	Lower 95%	Upper 95%	Classification
Mersey	1	69.90	19	60.28	79.53	Outlier
Eastern	2	67.30	23	59.53	75.07	
West Midlands	3	67.00	12	56.06	77.93	
North Western	4	69.06	12	61.16	76.97	
Oxford	5	71.65	12	65.78	77.53	Below mean and 2 <sup>nd</sup> quartile
Yorkshire	6	74.44	39	69.55	79.32	
Eastern	7	74.50	60	69.96	79.04	Not outlier
Scotland (West)	8	68.33	7	55.78	80.88	Bottom quartile Not below mean Not outlier
Scotland (West)	9	71.05	10	57.35	84.76	
Oxford	10	70.85	16	61.19	80.52	
London	11	70.62	7	60.22	81.02	
Scotland (West)	12	66.01	5	51.33	80.70	

<sup>20</sup> Adjusted for method variance – see Chapter 2

The proportion of outlying indicators across all available specialties was calculated for each provider. Of the providers in this analysis, 34 per cent (120/355) had no indicator scores that were outliers (taken from the Provider profile report). Only 15 per cent (52/355) had 10 per cent or more of indicator scores across all their specialties with respondents that were outliers.

## Deanery quality improvement

All deaneries were asked to provide an action plan in response to the deanery specific data provided from the National Trainee Survey. The deaneries were given a proforma with headings that are standard in most action plans (see Appendix 2 for a blank proforma at [www.pmetb.org.uk/traineesurvey](http://www.pmetb.org.uk/traineesurvey)).

The deaneries were given three months to return an action plan to PMETB<sup>21</sup>. All deaneries achieved a response, many with a full and considered action plan. In Scotland, NHS National Education for Scotland (NES) provided a Scotland wide response and then, in two cases, a deanery specific overview. The responses were variable in style and content. This was expected at this early stage and with the challenges of effective dissemination from PMETB to the deaneries, and from the deaneries to the education providers (e.g. NHS trusts and boards).

The data received by deaneries were affected by the filter to keep trainees' identities anonymous, in that specialty groups at providers for which less than four responses were received were not included in the report. For some specialty groups at some providers, deaneries have noted that more information is required to link the trainees' responses to a specific clinical environment/department and thus write a detailed action plan. PMETB will explore whether some specialties (as opposed to specialty groups) can be reported for some larger providers instead of just across the deanery. However, it will not be possible to report on something such as a department within a provider that is not defined by a provider listed by the National Administrative Codes Service and a specialty or sub-specialty.

Equally, deaneries need to adapt and plan their responses. Several of the deaneries decided to document their actions in order to create a full action plan which they will send to PMETB at a later point – an 'action plan for action plans'.

All deaneries used the outlier identification detailed above as the rationale for inclusion of a provider within an action point. The action plans were very clear on the levels for responsibility and generally the headings do not appear to have caused problems. Some of the deaneries were notable for their clarity of both responsibility and timescales. Virtually all deaneries evidenced a clear grasp of priorities and risk management, and there was significant consistency in the identification of high risk and appropriate timescales. One deanery used a three level approach, with urgent action within one month as the first level, the second level for action within three months and the remainder as developmental changes.

Several of the deaneries documented their concerns about disseminating the data in a useful way to their local education providers and PMETB will explore this further when consulting on reporting (see Chapter 5). Half of the deaneries felt able to be specific in identifying the education providers concerned. Several of the deaneries referred very helpfully to previous quality management activity and/or that the data matched their experience of the quality of education for those education providers. Occasionally, the deaneries have questioned a mismatch between their own and the education provider's experiences and the outcomes in specific areas such as handover. In addition to this retrospective analysis, several deaneries noted future activity which could occur where there was evidence of

<sup>21</sup> Proforma sent out 7 December 2006 with specialty group data by provider benchmarked to national means and quartiles (see table 21)

continuing problems. One deanery noted that the next step could be initiation of a triggered PMETB visit for one education provider.

The utility of specialty group based data and identification of specialty groups themselves was also interesting and will be considered in future development of the surveys. Some (less than 50 per cent) of the action plans identified the specialty group clearly and specifically. Other action plans did not identify the provider specialty group clearly, concentrating on the issue and the education provider(s). Several linked specialty group, the issue highlighted by the indicator score and the provider; this approach makes the action plan both understandable and easier to follow up. Once action planning is embedded in the deanery systems, the involvement of specialties and schools will become the norm. One deanery had very clear identification of the role of the schools and Specialist Training Committee chairs within the action plan; this was one of the more successful in integrating issue, specialty and provider.

Deaneries will use the action plans for several purposes, including deanery quality management, but also to provide the information and reassurance that PMETB will need that *Generic standards for training* are being attained and maintained. Form needs to follow function; it may be that action plans at a higher level and with broader scope will meet regulatory and performance management issues, while a more detailed approach is needed at quality management (deanery) and quality control (provider) levels. A lack of detail in many of the action plans means that external readers cannot assess accuracy, the relevance of issues or appropriateness of actions taken; neither can they identify if the actions have been undertaken.

Many of the deaneries and NES confirmed their intention to improve both the response rate and the accuracy of their population data (an issue discussed in Chapter 5).

## 5. The future of the National Trainee Survey

The National Trainee Survey of 2006 was the first national survey conducted by PMETB and COPMeD and there were significant challenges in its delivery and a number of important lessons learnt. The high response rate and the debate and discussion that followed the release of information so far indicates that there is a significant role for this survey in the development and quality assurance of postgraduate medical education in the future.

A number of changes are planned for the 2007 survey; these are described below and are based on PMETB's experience of the 2006 survey and feedback from the deanery staff that kindly helped with the work.

### Route of administration

To eliminate the method variance problem outlined in Chapter 2, where trainees' responses to some items varied by the route by which the survey was administered, the 2007 survey will only be administered using one route: a national website. Web administration of the survey has a number of benefits over the other methods used for the 2006 survey. These are:

1. It is more economical than scanning paper forms and the portable electronic survey units.
2. It allows the easy administration of specialty specific versions of the questionnaire as items can be conditionally displayed based on the responses to earlier items.
3. The system can produce a reference number so that doctors can prove they have completed the survey, thereby providing deaneries with the data to enforce the mandatory status of the future trainee surveys.
4. A web based survey will dramatically reduce local administrative burden. The task managed locally will be the collection of valid email addresses (which will have considerable collateral benefit).

It is known that trainees may alter their responses when they feel these could be identified (although assurances were given that the data were confidential and reporting would ensure that responses remained anonymous); this might reduce the total variance available for analysis from some items. However, this will not invalidate comparison between training providers, as it may be assumed that any measure of it is not associated with any particular provider since all are using the same method of administration and impression management<sup>22</sup> is an individual-level variable. There may be inter-specialty differences in impression management but, as benchmarking is within specialty group, any such differences will not bias the analysis.

### The content of the survey

Changes to the content of the survey are constrained by two factors:

1. The data will be used to test for changes over time; therefore the items used to derive the indicator scores must remain the same.
2. Shorter questionnaires are known to elicit a higher response rate<sup>xx</sup> and thus the questionnaire should not be substantially lengthened.

Within these constraints a number of changes are proposed for the 2007 work.

<sup>22</sup> Impression management: the extent to which an individual answers items to give a more positive impression of themselves.

## Adequate experience

The *sine qua non* of any post is to provide the trainee with the experience s/he needs to acquire the competences set out in the curriculum that s/he is following. Currently, this standard is addressed through two summative evaluation items on which trainees are asked to rate the quality of the experience overall. It has not been possible to demonstrate the construct validity of these items; it cannot be shown using factor analysis that the items capture something different from overall satisfaction with the post – a failure of discriminant validity. This means that the score is not adding substantial information to that which is already contained in the Overall Satisfaction Score. Furthermore, it does not provide diagnostic information on which aspects of the experience offered by the post were found to be inadequate. As all trainees received the same items in 2006, it would not have been possible to have items that related to facets of training experience, as these are specialty specific. However, in 2007, it will be possible to display items conditionally, based on the respondent's earlier answers. It will therefore be possible to ask about particular facets of experience in relation to both the trainee's specialty and stage of training. Item development for these speciality-specific adequate experience items will be done in conjunction with lead deans and the Royal Colleges and faculties. Items will be related to the PMETB approved curricula. The challenge will be to devise items that will provide sufficient data variability to distinguish between posts providing experience that trainees perceive as 'good' and posts that are not perceived as 'good'. Posts are part of a programme that is designed to provide the necessary experience overall, and thus it may be necessary to ask whether the post provided the experience the trainee expected, given the design of the overall programme.

## Overall Satisfaction Score

There is a ceiling effect on the Overall Satisfaction Score, which means it is not possible to distinguish between posts at the top end of the score. All doctors who are satisfied with their posts are alike as far as the survey data is concerned, 12.4 per cent (N = 24,848) of respondents having the maximum possible score on this scale. This measurement issue will be addressed in the 2007 survey with the addition of items written with a view to making distinctions possible at the positive end of the overall satisfaction score.

## Follow up of concerns

The survey currently promises confidentiality for respondents, but this promise conflicts with some respondents' expressed desire for action to be taken in response to concerns they have raised in free text comments. Therefore, the 2007 survey will allow respondents to indicate that they would be happy to be identified in the course of any follow-up action in response to their concerns. This approach is taken by the Training and Development Agency for Schools in their newly qualified teacher survey, which also examines the perceived quality of training received<sup>xxi</sup>.

## Personality variables

From the occupational psychology literature it is known that a respondent's personality accounts for some of the variance in their self-report of constructs such as job satisfaction<sup>xxii</sup>. Incorporating a personality measure in the model can improve the prediction of turnover<sup>xxiii</sup>; it might therefore be expected that incorporating a personality measure would improve the trainee survey's predictive validity (as the

effects of the personality measure can be statistically adjusted). Negative ratings from trainees who are generally positive but negative about their current training post are more likely to relate to problems with the posts than negative ratings from trainees who are generally negative. The 2007 survey will therefore include a shortened measure of positive and negative affect to allow for statistical adjustment of personality factors when using the Overall Satisfaction Score to benchmark training providers<sup>xxiv</sup>.

Responses varied by route of administration. For some items, more positive responses were more likely with less anonymous routes of survey administration and thus it would seem appropriate to include some items to measure the extent to which trainees are considering the impression of themselves that they are presenting when answering. This measure of impression management can then be tested for associations with items used in the indicator scores and, if necessary, statistically controlled for. Merrill *et al*<sup>xxv</sup> have developed a scale designed to measure impression management that is domain specific for medicine, which could be used by the 2007 survey.

## Dissemination of the results

The time from data collection to reporting of the 2006 data was longer than it should have been. Reporting was impeded by the following factors, all of which will not be present for the 2007 work:

1. Four routes of administration, each with variance from the master data template, had to be merged to create one data set.
2. Indicator scores had to be adjusted for method variance to allow comparison across training providers because the provider was confounded by the route of administration.
3. Some respondents' data had to be remapped to an alternative location. For instance, in one English Postgraduate Medical Education Centre, GPRs had received paper forms with the acute trust identifier.

It should therefore be possible to report the findings back to the deaneries and other stakeholders in a more timely fashion. A variety of data aggregations are appropriate, given the local systems in place for quality managing postgraduate medical education; these are described below in Table 23. Each aggregation reports on the same indicator scores. Aggregations are filtered, in that only providers or rotations with more than three respondents are reported to ensure that individual trainees' responses remain anonymous.

**Table 23 Aggregations for reporting**

Audience	Aggregation
Deans and Directors of Medical Education	Training provider and specialty group, for example anaesthetics trainees at St Elsewhere NHS Trust
Deans and Specialist Training Committees	Deanery Specialty Grade, for example London Cardiology SpR trainees

The Deanery Specialty aggregation ensures that the responses from trainees in the smaller specialties are reported on, as frequently there would not be enough responses to include this in the Training Provider and Specialty Group aggregation. For example, with 2006 data, occupational medicine was reported by Deanery Specialty Grade aggregation only and not by location, as there are not enough trainees at a given location to get above the reporting filter.

As the focus of the work is on quality management, reporting back to those with accountability for the training and seeking their responses to the data should always be the priority. However, other parties, not least the trainees who completed the survey, have a right to see the data for the locations of interest to them. To facilitate this, PMETB plans to report local findings through a web based reporting service. This will supplement or replace (depending on whether all the required functionality can be obtained) the existing method which uses a CD-ROM and the *Compare* software provided by the Healthcare Commission. PMETB will consult on the reporting options for the 2007 survey as part of the quality assurance (QA) framework consultation to establish:

- that the aggregations are at a level of granularity that allows data to be linked back to a clinical area for which an action plan can be written; constraints include the identification of the entity to aggregate to and the need to ensure that no one trainee's responses can be identified in the report;
- that the reporting tool allows deaneries to disseminate the findings to their providers with their local contextual information included;
- which aggregations are most important and should be reported on first;
- other desirable aggregations such as hospital based VTS GP trainees only;
- benchmarking groups – for each aggregation the units can be grouped in different ways for comparison purposes, for instance one could compare cardiology trainees' experience with the experience of all other medical specialties, or just the trainees in the acute medical specialties.

When considering proposals for reporting it is important to remember that PMETB has made the following guarantees:

- all survey data are confidential and it will not be possible for deanery staff to identify the responses of an individual trainee doctor;
- data will be reported in such a way that responses remain anonymous and cannot be identified due to small numbers in any particular group.

## Population data

The 2006 survey's population comprised all trainees in approved training posts. PMETB was dependent on deaneries to identify these doctors, as there is presently no single, centrally maintained UK database of trainees in approved posts. Headcount of doctors in approved posts by specialty, provider and grade was not available at the start of the data collection period due to time pressure, which meant that response rates could not be calculated immediately. In addition, there were issues around how posts are defined as approved, since there are posts that are deanery funded and educationally approved, posts that are locally funded and educationally approved, and posts that are service only. Information on these posts was not generally held in one place; for instance, deaneries hold SpR data since they issue National Training Numbers, but deaneries only hold whole time equivalent data for SHO posts that are deanery funded. Data quality issues were apparent; for instance, over 200 of the same SpRs were listed on two different deaneries' databases.

Both PMETB and the deaneries are committed to improving these data for the 2007 survey. PMETB has been consulting with deaneries' business managers over how best to obtain data on the population of trainees in approved posts, and will be issuing their data request well in advance of the 2007 survey. At a national and UK level, PMETB is working with representative groups to discuss the development of a minimum data set, of which these data will be part. These proposals will be included in the consultation.

Deaneries will know how many specialty trainee posts there are, as these were counted as part of the preparation for the implementation of Modernising Medical Careers (MMC)<sup>23</sup>. However, the national counting of fixed-term specialty training appointment (FTSTA) posts may be an issue as, according to the forthcoming *A guide to specialty training: the gold guide*<sup>xxvi</sup>, there is no mandatory requirement for deaneries to record these posts and currently no specified data template on which to do so. PMETB will consult with deaneries to establish an agreed template for recording details of these posts as part of the QA framework.

In England, the Department of Health conducts an annual census of medical staff<sup>xxvii</sup>. It uses pay scales to classify doctors by grade and does not collect data on whether the doctor is in an educationally approved post. Trainees are currently classified as:

- registrar group, which refers to the combined grouping of specialist registrars, senior registrars and registrars and other staff working at equivalent grades that are not in an educationally approved post;
- doctors in training and equivalents (previously known as junior doctors): registrar group, senior house officer, foundation doctors, house officers and other staff working at equivalent grades that are not in an educationally approved post.

PMETB has requested that the Information Centre for Health and Social Care in England include a field on the census to indicate whether the doctor is in an educationally approved post or not. PMETB will also establish whether the other three nations' departments conduct similar censuses of medical staff.

## **The survey as screening tool**

The utility of the survey as a screening tool is determined by its predictive validity - whether it can identify posts or programmes that may require remedial action or highlight areas of good practice. The forthcoming consultation on the QA framework will identify a criterion score that could be used to measure the survey's predictive validity. A measuring instrument may have to be developed; scores from this can be used as the dependent variable, with the indicator scores from the survey as the independent variables. This will establish whether trainees' perceptions of a provider correlate with the criterion score. To avoid range restriction, the criterion score will have to be collected from providers at all levels of the trainees' ratings, some from the bottom quartile and some in the middle, and so on for the duration of the exercise. For the purposes of the validation exercise, anyone involved in obtaining the criterion score will need to be blinded to the survey data.

The exercise could establish which indicator scores are most effective and adjust the identification of outliers to ensure that the survey is more sensitive than specific, as it may be better to have false positives than for PMETB or the deaneries to fail to follow up a problem.

<sup>23</sup> See: <http://www.mmc.nhs.uk/pages/home>

## **Making the survey mandatory**

Feedback from deanery and postgraduate centre staff indicates that making the survey mandatory could make administration far easier. Due to the different routes of administration employed, it would not have been possible to enforce this for the 2006 survey. The forthcoming *A guide to specialty training: the gold guide* and the General Medical Council's *Good medical practice*<sup>xxviii</sup> both indicate that doctors are required to take part in systems of quality assurance and quality improvement. The survey forms a key part of the quality assurance of doctors' training programmes and they will therefore be expected to complete a return for future surveys. As the survey will be exclusively administered through a website, which will provide reference numbers for successfully completed submissions, doctors will be able to show that they have completed the survey. In addition, PMETB will be able to list who has not completed a return and supply this information to the deaneries and centres for follow-up action. Together, this will provide a mechanism for deaneries to encourage a high response rate.

Recognising that PMETB is committed to the principles of better regulation and so has a duty to minimise the burden of its data collection activities, PMETB will ensure that doctors' participation in the work is as easy as possible and will therefore provide support throughout the survey period to ensure that doctors are able to complete a return. In addition, a full pilot of the process will be conducted in one deanery to spot any process issues prior to national roll-out.

## **Comparison with 2006 data**

Once data from two years are available it will be possible to perform analysis looking at the relationship between the 2006 and the 2007 indicator scores for particular aggregations; for instance, for a given specialty group, do the same locations come up as low scoring? In addition, the 2006 data provide a baseline prior to the implementation of MMC.

## **National use of the data by Royal Colleges**

For the 2006 survey, PMETB supplied Royal Colleges with national level aggregations for each specialty. These data will be made available using the 2007 data too, but for 2007 these data will be of more interest as they will include the specialty specific items that the colleges will have helped to develop.

## **The 2007 survey and other data sources**

The survey data can be used to test other hypotheses. For example, if it is known that some trusts are participating in a quality improvement activity and others within the deanery are not, a comparison of the relevant indicator score can be made between the two groups. Analyses such as these can be performed, providing deaneries submit the grouping data available and it is possible to specify which indicator scores should vary by group to avoid multiple testing.

During the 2007 trainee survey, a survey of trainers will be conducted concurrently. The aggregated data from these two surveys will be linked using specialty and location. This will allow comparison of the trainees' and trainers' perspectives and the potential development of discrepancy indicators; the greater the difference between the two perspectives, the greater the potential problem. It is likely that trainees' perceptions will not be as positive as those of trainers. Baker and Sprackling<sup>xxix</sup> compared the two perspectives (consultants and their VTS SHOs) and

found that consultants reported more positive perceptions. For instance, 32 per cent of the trainees in their sample indicated that teaching took place in protected time, whereas the figure for consultants was 67 per cent.

## Deanery action plans

To ensure that survey data are always used for quality improvement work, action plans should be written for an external audience so that trainees can see that their participation in the survey is a worthwhile activity, and that patients can be reassured that deaneries are effective in ensuring the safe training of doctors. PMETB will also be seeking evidence of trainee involvement in the agreement and measurement of the action plans.

## Indicative plan for the 2007 survey

Table 24 shows key external dates for the trainee in the 2007 survey plan. PMETB recognises that deaneries need sufficient notice that data will be released to allow them to schedule resources for onward dissemination of the data to their providers and the preparation of action plans in conjunction with the providers. PMETB will provide more detailed plans nearer the time.

**Table 24 2007 survey plan**

Milestone	Date
Development of specialty specific items in conjunction with lead deans and Royal Colleges/faculties.	April – July 2007
Population request to deaneries for completion. This will consist of list of approved posts detailing the specialty, grade and location of each.	June 2007
Population request returned to deaneries for them to supply the email addresses of the posts' current incumbents. A draft version of this request is available at <a href="http://www.pmetb.org.uk/traineesurvey">www.pmetb.org.uk/traineesurvey</a> .	August 2007
Full pilot of trainee questionnaire in one deanery.	October 2007
Staged launch of National Trainee Survey.	December 2007– January 2008
National Trainee Survey reports (all aggregations, see Table 22) made available to deaneries, Royal Colleges, faculties and specialist associations.	March 2008

## Feedback form

If you have comments on the 2006 work or the plans for the 2007 work, please go to [www.pmetb.org.uk/traineesurvey](http://www.pmetb.org.uk/traineesurvey) and complete the feedback form.

## References

- i PMETB. *Generic standards for training* (2006). Available at: [http://www.pmetb.org.uk/fileadmin/user/Policy/Policy\\_Statements/Generic\\_standards\\_for\\_training\\_April\\_06.pdf](http://www.pmetb.org.uk/fileadmin/user/Policy/Policy_Statements/Generic_standards_for_training_April_06.pdf)
- ii Grant, J, Maxted M, Owen H, Wooding K, Black C, Neville E, Gale R. *Development Of A National Generic Data Collection System To Support Inspection Of Hospital Visiting* (2004). Available at: <http://www.pmetb.org.uk/index.php?id=backgroundinfo#c977>
- iii Paice E, Aitken M, Cowan G and Heard S. (2000). Trainee satisfaction before and after the Calman reforms of specialist training: questionnaire survey. *British Medical Journal*;320: 832-836.
- iv Leach C and Goodman J. (2006). *PMETB Questionnaire Assessment Final Report*. Available at: <http://www.pmetb.org.uk/index.php?id=backgroundinfo#c977>
- v Baldwin D C Jr and Daugherty S R. Sleep deprivation and fatigue in residency training: results of a national survey of first- and second-year residents. *Sleep*. 2004 Mar 15;27(2):217-23.
- vi Healthcare Commission (2006). *Working in partnership: Getting the best from inspection, audit, review and regulation of health and social care*. Available here: [http://www.concordat.org.uk/\\_db/\\_documents/Concordat\\_-\\_version\\_\(May\\_06\).pdf](http://www.concordat.org.uk/_db/_documents/Concordat_-_version_(May_06).pdf)
- vii Bland M J, Altman D. (1995). Multiple significance tests: the Bonferroni Method. *British Medical Journal*. 1995;310:170.
- viii Nunally J C. (1978). *Psychometric theory* (2nd edn). New York: McGraw-Hill.
- ix Cronbach L J and Meehl P E. (1955). Construct Validity In Psychological Tests. *Psychological Bulletin*, 52, 281-302.
- x Cohen J and Cohen P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. (2nd ed.). Lawrence Erlbaum Assoc.
- xi Goldstein H. (1995). *Multilevel statistical models*. 2 ed. London: Arnold.
- xii McManus C. (2007). *Multilevel modelling of PMETB data on trainee satisfaction and supervision*. Report produced for PMETB available at: <http://www.pmetb.org.uk/index.php?id=2006results#c978>
- xiii Hicks B. (2000). Time to stop bullying and intimidation. *Hospital Medicine*;61:428-31.
- xiv Paice E, Aitken M, Houghton A, Firth-Cozens J. (2004). Bullying among doctors in training: cross sectional questionnaire survey. *British Medical Journal*;329(7467):658-9.
- xv Altman D, Machin D, Bryant T N and Gardner M J. (2000). *Statistics With Confidence: Confidence Intervals and Statistical Guidelines*. London: BMJ Books
- xvi Frank E, Carrera J S, Stratton T, Bickel J and Nora L M. Experiences of belittlement and harassment and their correlates among medical students in the United States: longitudinal survey. *British Medical Journal*;333:682-8.
- xvii For details see: <http://www.riti.org.uk/index5.html>

- xviii For details see:  
[http://www.healthcareworkforce.nhs.uk/index.php?option=com\\_d4j\\_ezine&Itemid=332](http://www.healthcareworkforce.nhs.uk/index.php?option=com_d4j_ezine&Itemid=332)
- xix Mahon A, Harris C, Tyrer J, Carr S, Lawson K, Carr L, Chaplin S and Wright D. (2005). *The implementation and impact of Hospital at Night pilot projects: An evaluation report*. Department of Health. Available at:  
<http://www.wise.nhs.uk/sites/workforce/usingstaffskillseffectively/Hospital%20at%20Night%20Document%20Library/1/Hosp%20At%20Night%20Final%20Report.pdf>
- xx Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I. (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*;324;1183-1191.
- xxi Training and Development Agency for Schools (2006). *Results of the newly qualified teacher survey 2006*. Available here:  
<http://www.tda.gov.uk/upload/resources/pdf/n/nqt%20survey%202006%20-%20sector%20level%20report.pdf>
- xxii Spector P E. (1994). Using self-report questionnaires in OB research: a comment on the use of a controversial method. *Journal of Organizational Behaviour*, 15, 385-392.
- xxiii Judge T A. (1993). Does Affective Disposition Moderate the Relationship Between Job Satisfaction and Voluntary Turnover. *Journal of Applied Psychology*, 78, 3, 395-401.
- xxiv Mackinnon A, Jorm A F, Christensen H, Korten A E, Jacomb P A and Rodgers N. (1999). A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences*, 27, 405-416.
- xxv Merrill J M, Laux L F, Lorimor R J, Thornby J I, Vallbona C. (1995). Measuring social desirability among senior medical students. *Psychological Reports*. 1995 Dec;77(3 Pt 1):859-64.
- xxvi Department of Health (In press 2007). *A guide to specialty training: the gold guide*.
- xxvii The Information Centre (April 2006). *Hospital and Community Health Services Medical and Dental staff in England: 1995-2005*. Available here:  
<http://www.ic.nhs.uk/pubs/nhsstaff/medicaldentalintro/file>
- xxviii General Medical Council (2006). *Good Medical Practice*. Available here:  
[http://www.gmc-uk.org/guidance/good\\_medical\\_practice/GMC\\_GMP.pdf](http://www.gmc-uk.org/guidance/good_medical_practice/GMC_GMP.pdf)
- xxix Baker M and Sprackling P D. (1994). The educational component of senior house officer posts: differences in the perceptions of consultants and junior doctors. *Postgraduate Medical Journal*. 1994 Mar;70(821):198-202.



---

---

---

Postgraduate Medical Education  
and Training Board

Hercules House  
Hercules Road  
London SE1 7DU

Tel +44 (0)20 7160 6100

Fax +44 (0)20 7160 6102

[www.pmetb.org.uk](http://www.pmetb.org.uk)

© PMETB 2007

ISBN: 978-0-9555910-0-6

