Acknowledgements

The project described in this report has involved over 1500 people, most of whom have given their time on a voluntary basis. The project team would particularly like to thank:

- The Department of Health, for commissioning the project
- The members of the Project Board and Project Group, for their advice and guidance
- The volunteers from across the UK who have written SJT items
- The foundation doctors and concordance panel members who have reviewed SJT items
- Medical school staff who have helped develop the EPM, and have organised and run the pilot SJTs
- The students who gave up their time to sit the SJT
- The student representatives at medical schools who helped to promote the pilots
- The BMA Medical Students Committee, for their support
- The UK Foundation Programme Office, and the Rules Group, for their advice and support
- The NHS patients, nurses, and doctors who helped with the Job Analysis
- Our specialist advisors and suppliers

As Chair of the ISFP Project Group I should like to record formally Siobhan Fitzpatrick’s unstinting dedication to this project and to thank her for her exceptional contribution. Ably assisted by Amy Stringer, and latterly, Oliver Watson, the team has worked tirelessly to deliver this project on time – and within budget – thus permitting the forthcoming Parallel Recruitment Exercise to be undertaken.

Professor Paul O’Neill
Chair, ISFP Project Group
Table of Contents

LIST OF FIGURES ................................................................................................................................. 4
LIST OF APPENDICES .......................................................................................................................... 4
GLOSSARY .................................................................................................................................................. 5

1. EXECUTIVE SUMMARY .................................................................................................................. 6
2. INTRODUCTION ............................................................................................................................... 7
3. BACKGROUND ................................................................................................................................. 7
   3.1. Existing arrangements for selection to the Foundation Programme .............................................. 8
   3.2. The case for change ...................................................................................................................... 9
4. ISFP PROJECT PLAN ....................................................................................................................... 12
   4.1. Objectives .................................................................................................................................. 12
   4.2. Approach and timescales ............................................................................................................. 12
   4.3. Management and Governance ................................................................................................... 13
5. JOB ANALYSIS OF THE ROLE OF THE FY1 DOCTOR ................................................................ 14
   5.1. Methods .................................................................................................................................... 14
   5.2. Professional Attribute Framework .............................................................................................. 15
6. SJT ITEM WRITING AND QA PROCESS ..................................................................................... 17
   6.1. SJT design ................................................................................................................................. 17
   6.2. Item development and review ..................................................................................................... 18
   6.3. Item writing workshops .............................................................................................................. 19
   6.4. Initial item review ....................................................................................................................... 20
   6.5. Lead item review ......................................................................................................................... 21
   6.6. Foundation doctor focus groups ................................................................................................ 21
   6.7. Concordance Panel .................................................................................................................... 22
   6.8. Test construction ......................................................................................................................... 22
   6.9. Scoring convention ..................................................................................................................... 23
7. SJT PILOTS ......................................................................................................................................... 24
   7.1. Overview .................................................................................................................................... 24
   7.2. Methods of SJT delivery ............................................................................................................. 24
   7.3. Administration ............................................................................................................................ 28
   7.4. Participation ............................................................................................................................... 31
   7.5. Feedback from pilots ................................................................................................................. 31
   7.6. Initial pilot - key findings ........................................................................................................... 32
   7.7. Large-scale pilot - key findings .................................................................................................. 33
8. EDUCATIONAL PERFORMANCE MEASURE (EPM) .................................................................... 34
   8.1. Background ............................................................................................................................... 34
   8.2. Consultation and piloting ............................................................................................................ 35
   8.3. Pilot of draft EPM framework .................................................................................................... 36
   8.4. EPM pilot lessons ...................................................................................................................... 38
   8.5. EPM Task and Finish Group (EPM Group) ............................................................................... 38
   8.6. Draft framework ....................................................................................................................... 38
   8.7. Performance in relation to graduating cohort ........................................................................... 39
   8.8. Degrees, presentations, prizes and publications ...................................................................... 39
   8.9. Extra-curricular activities ......................................................................................................... 39
9. ASPECTS OF THE APPLICATION PROCESS USING SJT AND EPM .................................... 40
   9.1. Aggregating the SJT and EPM scores ....................................................................................... 40
   9.2. Algorithm .................................................................................................................................. 40
   9.3. Development and testing of SJT item banking software .......................................................... 41

ISFP Final Report  Page 2
List of Figures

Figure 1: Project Plan .............................................................................................................................. 13
Figure 2: Governance structure for ISFP project .................................................................................... 13
Figure 3: Membership of the ISFP Project Board .................................................................................. 13
Figure 4: Membership of the ISFP Project Group .................................................................................. 14
Figure 5: Professional Attributes Framework .......................................................................................... 16
Figure 6: Relationship between Professional Attributes and SJT target domains ..................................... 17
Figure 7: Process of SJT item development and review ........................................................................ 19
Figure 8: SJT item writer demographics and job roles .......................................................................... 20
Figure 9: Total number of SJT items written by target domain .............................................................. 20
Figure 10: Focus group participant demographics .................................................................................. 21
Figure 11: Scoring key for SJT ranking items ......................................................................................... 23
Figure 12: Methods of SJT test delivery .................................................................................................. 27
Figure 13: Administrative lessons for SJT delivery ................................................................................. 30
Figure 14: SJT pilot participant demographics (initial pilot, Autumn 2010) .............................................. 32
Figure 15: SJT pilot participant demographics (large-scale pilot, Spring 2011) ...................................... 33
Figure 16: Participation in three papers of large-scale pilot .................................................................. 33
Figure 17: Alternative selection tools considered during Options Appraisal ........................................ 51
Figure 18: Relative costs of selection tools in Options Appraisal .......................................................... 51
Figure 19: Criterion for non-financial costs and benefits ....................................................................... 52
Figure 20: Non-financial cost benefit analysis ....................................................................................... 52
Figure 21: The case for change revisited ............................................................................................... 54

List of Appendices

Appendix A: Legal opinion, July 2011 ................................................................................................. 59
Appendix B: Characteristics of the current system for selection to the Foundation Programme .......... 71
Appendix C: Project Initiation Document, February 2010 ................................................................. 77
Appendix D: FY1 Job Analysis ............................................................................................................. 125
Appendix E: Person Specification FP2012 ........................................................................................... 241
Appendix F: Report of the SJT Pilots .................................................................................................... 245
Appendix G: SJT Pilot Administrators Guidance ................................................................................... 375
Appendix H: Report of EPM framework pilot ....................................................................................... 389
Appendix I: Report of the EPM Task and Finish Group ..................................................................... 417
Appendix J: Report of aggregating SJT and EPM scores .................................................................. 435
Appendix K: ISFP Communications Plan ........................................................................................... 445
### Glossary

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>Academic Foundation Programme</td>
</tr>
<tr>
<td>AoMRC</td>
<td>Academy of Medical Royal Colleges</td>
</tr>
<tr>
<td>BMA</td>
<td>British Medical Association</td>
</tr>
<tr>
<td>BMA MSC</td>
<td>British Medical Association Medical Students Committee</td>
</tr>
<tr>
<td>BME</td>
<td>Black and Minority Ethnic</td>
</tr>
<tr>
<td>BMJ</td>
<td>British Medical Journal</td>
</tr>
<tr>
<td>CBA</td>
<td>Cost Benefit Analysis</td>
</tr>
<tr>
<td>COPMeD</td>
<td>Conference of Postgraduate Medical Deans</td>
</tr>
<tr>
<td>CPD</td>
<td>Continuing Professional Development</td>
</tr>
<tr>
<td>DH</td>
<td>Department of Health</td>
</tr>
<tr>
<td>DPMD</td>
<td>Defence Postgraduate Medical Deanery</td>
</tr>
<tr>
<td>EEA</td>
<td>European Economic Area</td>
</tr>
<tr>
<td>EMQ</td>
<td>Extended Matching Question</td>
</tr>
<tr>
<td>EPM</td>
<td>Educational Performance Measure</td>
</tr>
<tr>
<td>FP</td>
<td>Foundation Programme</td>
</tr>
<tr>
<td>FPAS</td>
<td>Foundation Programme Application System</td>
</tr>
<tr>
<td>FY1/FY2</td>
<td>First/ Second year of an integrated two year Foundation Programme</td>
</tr>
<tr>
<td>GMC</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>GP</td>
<td>General Practice</td>
</tr>
<tr>
<td>ISFP</td>
<td>Improving Selection to the Foundation Programme project</td>
</tr>
<tr>
<td>MCQ</td>
<td>Multiple Choice Question</td>
</tr>
<tr>
<td>MDAP</td>
<td>Multi-Deanery Application Process</td>
</tr>
<tr>
<td>MEE</td>
<td>Medical Education England</td>
</tr>
<tr>
<td>Mini-CEX</td>
<td>Mini Clinical Evaluation Exercise</td>
</tr>
<tr>
<td>MMI</td>
<td>Multiple Mini Interview</td>
</tr>
<tr>
<td>MoD</td>
<td>Ministry of Defence</td>
</tr>
<tr>
<td>MSC</td>
<td>Medical Schools Council</td>
</tr>
<tr>
<td>MSC-AA</td>
<td>Medical Schools Council Assessment Alliance</td>
</tr>
<tr>
<td>MTAS</td>
<td>Medical Training Application System</td>
</tr>
<tr>
<td>NACT</td>
<td>National Association of Clinical Tutors</td>
</tr>
<tr>
<td>NEAF</td>
<td>National Education Advisors Forum</td>
</tr>
<tr>
<td>NEO PI-R</td>
<td>Psychological personality measure</td>
</tr>
<tr>
<td>NES</td>
<td>NHS Education for Scotland</td>
</tr>
<tr>
<td>NHSE</td>
<td>NHS Employers</td>
</tr>
<tr>
<td>NIMDTA</td>
<td>Northern Ireland Medical and Dental Training Agency</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Examination</td>
</tr>
<tr>
<td>PID</td>
<td>Project Initiation Document</td>
</tr>
<tr>
<td>PLAB</td>
<td>Professional and Linguistic Assessments Board</td>
</tr>
<tr>
<td>PRE</td>
<td>Parallel Recruitment Exercise</td>
</tr>
<tr>
<td>Prince 2</td>
<td>Method of project management</td>
</tr>
<tr>
<td>QABME</td>
<td>Quality Assurance of Basic Medical Education</td>
</tr>
<tr>
<td>SBA</td>
<td>Single Best Answer</td>
</tr>
<tr>
<td>SJT</td>
<td>Situational Judgement Test</td>
</tr>
<tr>
<td>SSC</td>
<td>Student Selected Components</td>
</tr>
<tr>
<td>UKCDR</td>
<td>UK Collaboration for a Digital Repository</td>
</tr>
<tr>
<td>UKFPO</td>
<td>UK Foundation Programme Office</td>
</tr>
</tbody>
</table>
1. Executive Summary

Since 2006 there has been a UK-wide process for selecting graduating doctors to the integrated two year Foundation Programme training in the NHS. Eligible applicants are selected according to an academic quartile score reflecting educational performance in relation to the graduating medical school cohort; evidence of degrees, presentations, prizes and publications; and answers to ‘white space’ questions on an online application form.

In 2009, in response to concerns about the longevity of the current national application process, the Department of Health commissioned a review of selection to the Foundation Programme, with the aim of recommending a more reliable, robust, valid, feasible and sustainable method for selection which would minimise the risk of successful legal challenge. The ensuing work was a collaborative venture led by the Medical Schools Council (MSC), involving the Academy of Medical Royal Colleges (AoMRC), the British Medical Association (BMA) Medical Students Committee, the Conference of Postgraduate Medical Deans (COPMeD), the General Medical Council (GMC), The National Association of Clinical Tutors (NACT), the Northern Ireland Medical and Dental Training Agency (NIMDTA), NHS Employers, the Scottish Board for Academic Medicine, the Scottish Foundation Board, the UK Foundation Programme Office (UKFPO) and the four UK Health Departments.

The Options Appraisal involved a thorough and detailed appraisal of the methods that might be used for selecting applicants to the Foundation Programme. The Steering Group concluded that the most appropriate selection method would be an invigilated ‘Situational Judgement Test’ (SJT) to replace ‘white space’ questions, and a more granular and standardised ‘Educational Performance Measure’ (EPM) to replace the ‘academic quartile’ scores.

The DH accepted the recommendations that the SJT and EPM should be developed and piloted for use in the context of selection to the Foundation Programme. In January 2010, the DH commissioned the MSC to run an 18-month project to complete the development and piloting. Representatives from the fifteen stakeholder organisations represented through the initial Options Appraisal convened as the ISFP Project Group, with some small changes in membership, to oversee the project. This document is the final report of the Improving Selection to the Foundation Programme project.

The key achievements of the project are as follows:

- An in-depth literature review on the use of SJTs for selection, particularly in the context of medicine
- An in-depth literature review of the methods of selection into medical education and training up to and including the Foundation Programme
- A formal Job Analysis of the role of the FY1 doctor, which identified nine professional attributes (Commitment to Professionalism, Coping with Pressure, Effective Communication, Learning and Professional Development, Organisation and Planning, Patient Focus, Problem Solving and Decision Making, Self Awareness and Insight and Working Effectively as Part of a Team) and 111 individual behaviours
- The production and evaluation of a design specification (number of items per SJT, duration of the SJT, scoring convention, etc) for a reliable and valid SJT, based on the target professional domains identified in the Job Analysis of the role of the FY1 doctor
- Volunteers from across the UK have been trained to write SJT items against the target professional domains (known as SJT ‘items’)
- A significant number of SJT items has been developed, quality assured and tested
- The SJT has been piloted by more than 1100 students from 15 UK and 2 non-UK medical schools
- The results of the SJT pilots have been subject to psychometric analysis, which has shown the SJT to be a reliable selection method in this context, capable of differentiating between applicants
- Standards for the delivery and administration of the SJT have been defined

---

1 Medical Schools Council (2009) Selection into the Foundation Programme: An Option Appraisal
Improving Selection to the Foundation Programme

The results of the work to develop and pilot new methods of selection demonstrate that the use of the SJT and EPM in combination, to replace the assessment methods currently used, will make selection to the Foundation Programme more reliable, robust and cost effective.

Summary of recommendations

- Selection to the Foundation Programme should reflect the skills, knowledge and professional behaviours of the applicant, reflecting the integrated nature of the Foundation Programme as both education and employment
- Selection to the Foundation Programme from FP2013 onwards should be based upon:
  - An invigilated Situational Judgement Test (SJT) to assess aptitude for the Foundation Programme (to replace ‘white space’ questions); and
  - An Educational Performance Measure (EPM) to reflect educational performance at medical school up to the point of application to the Foundation Programme (to replace quartiles)
- Other aspects of the process of application to the Foundation Programme should remain unchanged, namely a national application process including applicant declaration of foundation school preferences, the use of a matching algorithm and eligibility checking
- There should be a full-scale shadow Parallel Recruitment Exercise (PRE) SJT and EPM run concurrently with the current application process for FP2012
- Using data from tracking, there should be ongoing evaluation of the validity of the SJT and EPM as methods for selection to the Foundation Programme

2. Introduction

This document is the final report of the ISFP Project Group. The aim of the report is:

- To provide the UK health departments with detailed recommendations about whether and how SJTs and EPM scores should be introduced for the selection of applicants to the Foundation Programme; and
- To explain the work and evidence underpinning the recommendations

This report is not designed to be a guide for medical schools, students and others with an interest in selection to the Foundation Programme. Separate guidance will be produced if the recommendations in this report are accepted.

The report begins by describing the background to the project; it then summarises the work done and the findings, before presenting the recommendations. Supporting detail is provided in the appendices.

More information about the project can be found on the ISFP website at www.ISFP.org.uk

3. Background

This chapter summarises the background to the project to allow a general reader to appreciate the context of the report. It starts with an explanation of the existing arrangements for the selection of final year undergraduate medical
students from UK medical schools and eligible non-UK applicants to the Foundation Programme, and then summarises the case for change that has led to the project.

3.1. Existing arrangements for selection to the Foundation Programme

The UK Foundation Programme is a two-year structured integrated training programme for newly graduated doctors, designed to give trainees a range of experience in secondary and primary care environments before beginning specialty training. Successful completion of Foundation Year One (FY1) is a requirement for full GMC registration. The Foundation Programme represents a full-time period of education, training and employment, and selection to the Foundation Programme is thus subject to European employment law. Foundation Programme training is organised and delivered by foundation schools, who liaise with the NHS organisations in their area to ensure the availability of suitable training posts. Confirming the successful completion of the FY1 year is formally the responsibility of the universities that have medical schools under the Medical Act 1983 and its subsequent amendments, although in practice this is delegated to the local postgraduate dean.

Since 2005 there has been a UK-wide application process for selection to the Foundation Programme, consisting of a national, online application open to both UK and non-UK applicants who meet the eligibility criteria. Applicants are awarded an application score, comprised as follows:

- Academic quartiles to reflect educational performance at medical school (40%); and
- Additional academic achievements (degrees, presentations, prizes, publications) (10%); and
- Answers to free text ‘white space’ questions, which assess the national Person Specification (50).

The graduating medical school provides an academic quartile score (34, 36, 38 or 40 points) to reflect educational performance relative to all of the other students in the applicant’s year at that school. Applicants provide evidence of additional academic achievements, which are verified by a national panel against specific criteria. ‘White space’ questions and their scoring key are developed by the UKFPO Rules Group to assess the elements of the national Person Specification not relating to clinical knowledge and skills, with an extensive Quality Assurance process. The scoring of answers to the ‘white space’ questions is organised by foundation schools, and is undertaken by trained representatives - mainly clinicians - from the NHS.

The application score, derived by adding the scores for academic quartiles, additional academic achievements and ‘white space’ questions, are used to determine which applicants should be selected into the national Foundation Programme, and to allocate successful applicants to individual foundation schools. The allocation of applicants to foundation schools is determined by a set of rules (the allocation algorithm), which take into account the applicants’ scores and preferences in a particular way. If there are more applicants than posts, then the highest scoring applicants are selected. If there are more posts than applicants, then all eligible applicants are considered for allocation. Entry into the Foundation Programme is competitive. In 2010, about 7,700 students graduated from UK medical schools, and forecasts from the Medical Schools Council suggest that this number will rise to nearly 8,000 over the next two years. There were also 1,605 applications to the Foundation Programme from graduates of non-UK medical schools; of which 270 were eligible. In total there were 7,253 eligible applicants for 7,073 Foundation Programme places.

Selection to the Foundation Programme comprises three main stages: a) selecting the most suitable eligible applicants to the Foundation Programme, b) allocation to a specific foundation schools (national process) and c) matching to a specific programme within the foundation school (local process). This project to review the methods for national selection to the Foundation Programme relates to (a) and (b) only, but the processes involved with matching to

2 See www.foundationprogramme.nhs.uk
3 This is a simplification. Under the prevailing right-to-work legislation, applicants who have an unconditional right to work in the UK will be selected for places first, in order of score. Any unfilled places will then be allocated to the remaining applicants (e.g. those without an unconditional right to work in the UK), again in order of score.
individual programmes remains the responsibility of the relevant foundation schools, and is outside of the scope of the ISFP project. Further detail about the related characteristics of this matching process, including special circumstances and linked applications, are provided in Appendix B.

The annual application process, which is known as the Foundation Programme (FP) recruitment round, starts in the autumn of one year, the results are announced in the spring of the following year, and the applicants take up their posts at the beginning of August in that year. For most applicants, the process runs in parallel with their final year of undergraduate education.

The UK Foundation Programme Office (UKFPO) administers the recruitment rounds under contract to the four UK Health Departments. There is a separate contract with the Department of Health in England (DH) for the provision of the online system, known as the Foundation Programme Application System (FPAS), through which the applications, scoring and allocations are managed. A stakeholder Rules Group maintains the rules and standards for the application process. This includes setting the timetable for each round, agreeing the ‘white space’ questions and the associated scoring guidelines, quality assuring the scoring, and adjudicating over special cases.

The guiding principles and parameters for the Improving Selection to the Foundation Programme project are that:
- In order to protect patients and to reassure employers, it is necessary to ensure that all foundation doctors are fit for purpose, and, in the event of over subscription to the Foundation Programme, the most appropriate applicants are selected
- Selection to the Foundation Programme should be informed by a measure of clinical skills and knowledge
- Selection to the Foundation Programme should be informed by a measure of meeting the national Person Specification
- Performance in medical school should inform selection to the Foundation Programme
- Allocation to foundation programmes occurs before final year examinations have been taken, and so it is not possible to use performance in Finals as a selection tool

3.2. The case for change
Surveys conducted after each recruitment round have shown a growing level of satisfaction among applicants. However, there are shared anxieties amongst key stakeholders relating to the individual selection tools and the potential for the application process to be subject to legal challenge, as follows:
- The marking of the answers to the ‘white space’ questions is labour intensive, the cost approaching £2m per year in clinician time
- Applicants are allowed to complete the ‘white space’ questions online at their convenience. While this is logistically simple, it means that there is no real safeguard that the answers are the genuine sole work of the applicant
- Model answers to the ‘white space’ questions have been offered for sale over the internet. There are concerns that the impact of such model answers is likely to increase with time, as it will become increasingly difficult to write entirely fresh questions for each successive recruitment round, and the discrimination between applicants will diminish
- While the ‘white space’ questions appear to offer a practical way to rank large numbers of comparable applicants, and are derived from the national Person Specification, their technical reliability and validity is open to question
- There have been questions about the extent to which the academic quartile scores can be equated across medical schools, particularly since each school derives its scores as it sees fit rather than following a standard approach
Given these concerns, if there is an increase in the number of eligible non-UK applicants such that the Foundation Programme is over-subscribed, there are concerns that the selection process may be subject to legal challenge from unsuccessful applicants to the Foundation Programme.

To date, the impact of these concerns has been softened considerably by the fact that there have been fewer eligible applicants than Foundation Programme places up to the FP2011 recruitment round (meaning that all applicants can get a job) and historically around 90% of applicants have been allocated to their first preference foundation school. Hence, the process has generally been seen as positive by those involved. However, this perspective is somewhat misleading. The percentage of applicants who get their first choice is a product of the matching algorithm that is used and the relative demand for places at different foundation schools - it is not dependent on the application process or the scoring method. Furthermore, there is no guarantee that the number of posts will continue to exceed the number of applicants. The effect of European Directives on the Mutual Recognition of Professional Qualifications[^4] is that applicants from the European Union are entitled to compete equally for Foundation Programme training, provided that they meet the eligibility criteria; and in any event non-EU applicants can apply subject to the prevailing ‘right to work’ arrangements. In the event that the number of applicants exceeds the number of posts - as is likely in future - then it would be even more important that the application process ensures that the ‘best’ applicants are selected, and that the process is not open to likely successful legal challenge.

The 2008 Next Stage Review[^5] recognised that the current arrangements were not sufficiently robust, and recommended new work to develop a more reliable and valid selection process. As a result, the Department of Health (DH) commissioned the Medical Schools Council to set up a cross-stakeholder Steering Group to undertake an options appraisal for selection to the Foundation Programme, with the aim of recommending a more reliable, robust, valid, feasible and sustainable method for selection which would minimise the risk of successful legal challenge. The ensuing work was a collaborative venture between fifteen stakeholder organisations[^6], and reviewed evidence for the use of selection tools including personality questionnaires, structured interviews, unstructured interviews, multiple mini interviews, assessment centres, national examinations, portfolios, structured records of achievement, educational performance measurements, situational judgement tests, through:

- Surveys and consultations
- Meetings with stakeholders
- Three independent academic literature reviews
- A report from a panel of international experts in assessment and selection
- An independent cost-benefit analysis

The ISFP Steering Group submitted its final report of the Options Appraisal to the DH in September 2009, in which it made recommendations about the Foundation Programme application process. The relevant findings of the Options Appraisal are summarised as follows:

- The current selection tools, ‘white space’ questions and academic quartiles, are not sustainable for use in the long term, and the evidence of the Option Appraisal demonstrated a need for change
- As the Foundation Programme is an integrated training programme of education and employment, the selection of applicants should take into account their wider professional attributes for employment as well as their clinical knowledge and skills

[^4]: Article 53 of the Treaty on the Functioning of the EU and Directive 2005/36/EC
[^6]: The Academy of Medical Royal Colleges (AoMRC), the British Medical Association (BMA) Medical Students Committee, the Conference of Postgraduate Medical Deans (CoPMED), the General Medical Council (GMC), the Medical Schools Council (MSC), the National Association of Clinical Tutors (NACT), the Northern Ireland Medical and Dental Training Agency (NIMDTA), NHS Employers (NHSE), the Scottish Board for Academic Medicine, the Scottish Foundation Board, the UK Foundation Programme Office (UKFPO) and the four UK Health Departments.
• The three selection tools shortlisted for use as a measure of an applicant's wider professional attributes for employment, as defined by the national person specification, were i) structured interviews, ii) Multiple Mini Interviews (MMIs) and iii) a Situational Judgement Test (SJT)
  
• Whilst structured interviews are popular with prospective applicants, they would not be a robust basis for selection on the scale of selection to the Foundation Programme (n=8000), which has a common national Person Specification – interviews are more appropriate for selection to specific posts

• MMIs, in which each applicant undertakes several structured interviews with different interview panels, would be more reliable and more valid, but they would be labour intensive and expensive to implement and sustain

• The Options Appraisal concluded that an SJT of likely professional behaviours and judgements in hypothetical work situations would be the most appropriate measure of the professional attributes in the national Person Specification. As the SJT would be an invigilated, machine-markable test, it would be a secure, reliable, standardised and relatively inexpensive selection tool

• The two selection tools shortlisted for use as a measure of an applicant’s clinical knowledge and skills were i) a special selection test (a ‘national exam’ for ranking) and ii) a measure of educational performance at medical school based on existing information

• A national exam has the advantage of being easy to standardise, but would not be fair to all applicants owing to variations in the timing of curricula across medical schools and the timing of the application process (before finals). It would also be prohibitively expensive to develop, and was considered by key stakeholders as likely to undermine educational objectives through distracting applicant focus and distorting the delivery of the medical school curricula

• The Options Appraisal concluded that the considerable existing information available from medical schools should be used as a measure for applicant educational performance, given that allocation to posts must take place before the results of medical school Finals are known, as an Educational Performance Measure (EPM). In many ways the EPM would evolve from the academic quartiles, but it should be produced in a more robust, complete and standardised way, and possibly be more granular

• Work should begin to develop and pilot the SJT and EPM, with a view to implementation for selection to FP2013 onwards, when the contract for the current on-line FPAS system is due for renewal. Aside from testing the individual effectiveness of the SJT and EPM, the pilots would provide data that could be used to determine the relative weightings to be given to scores for the SJT and EPM components when they are combined to give an overall application score for the purpose of selection to the Foundation Programme

• An initial step in the work should be a formal Job Analysis of the role of the FY1 doctor in order that selection tools might be based on more detailed information about the role

• Other aspects of the process of application to the Foundation Programme should remain unchanged, namely a national application process including applicant declaration of preferences, the use of a matching algorithm and eligibility checking

• Work should continue outside of the scope of the ISFP project to develop common final examination questions between UK medical schools, and to develop an assessment of clinical communication for applicants without English as a first language and who have not interacted with patients in English during the course of their undergraduate medical degree

The DH accepted the recommendations of the ISFP Steering Group, and authorised a short feasibility study to determine how the recommendations might be taken forward. The feasibility study concluded that the recommendations should be implemented through two sequential projects. The first – the current project - would develop, refine and validate the specifications for SJT and EPM through pilots that would take place in parallel with the FP2011 recruitment round. Subject to ministerial approval of the outcome, a second project would then be commissioned to plan and prepare for implementation of the SJT and EPM in a new selection process to be used for live selection for the FP2013 recruitment round onwards
Following the report from the feasibility study, the DH authorised the Improving Selection to the Foundation Programme (ISFP) project, which began in January 2010.

4. ISFP Project Plan

The Improving Selection to the Foundation Programme (ISFP) project was set up in January 2010, following a short feasibility study in which a detailed Project Initiation Document (PID) had been produced. Key points from the PID, which is provided as Appendix C, are summarised in this chapter to provide the context against which the outcome of the project is described in subsequent chapters of this report.

4.1. Objectives

The main objective of the project was to design, develop, and pilot the SJT and EPM to replace the existing selection arrangements for use from the FP2013 recruitment round onwards. The scope of the project was to include:

- A detailed Job Analysis of the role of the FY1 doctor to identify the detailed criteria that should be used for the selection of applicants to the Foundation Programme using the SJT
- The development of specifications for the SJT and EPM
- The production of SJT items for use in live recruitment rounds
- The development or purchase of software to store SJT items securely
- The piloting of the SJT and EPM
- Collection of data about costs, resource requirements, timings, etc. for planning a live implementation of the new selection arrangements
- Developing expertise to underpin the success of the live implementation
- Raising awareness of the proposed changes to the selection process, and maintaining the support and cooperation of stakeholders
- Obtaining legal opinion to confirm that the SJT and EPM would be likely to withstand legal challenge

The live implementation of the SJT and EPM would be subject to confirmation of the associated business case and the DH approvals process.

4.2. Approach and timescales

The original plan for the project recognised that the SJT and EPM could be developed independently of each other. Of the two selection tools, the SJT represented by far the larger piece of work, and accounted for almost the entire critical path of the project. Given this, the project was managed in stages that broadly reflected the main steps in the development and piloting of the SJT, which were as follows:

- **Stage 1**: A Job Analysis and literature review to identify the professional attributes of the FY1 doctor which would be assessed by the SJT; the design specification of an SJT; the production and quality assurance of an initial set of SJT items
- **Stage 2**: An initial series of micro-pilots to try-out the SJT on a small scale
- **Stage 3**: The development of further SJT items for a large-scale pilot
- **Stage 4**: Analysis and reporting of outcomes of the large-scale pilot

The original project plan is illustrated in Figure 1.
4.3. Management and Governance

The management structure for the ISFP project is illustrated in Figure 2.

The ISFP Project Group, which was established to continue the work of the ISFP Steering Group during the Options Appraisal, reported to a Project Board established by the Department of Health. Formal highlight reports were produced each month. The membership of the ISFP Project Board is shown in Figure 3.

<table>
<thead>
<tr>
<th>Project Board Member</th>
<th>Representing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Patricia Hamilton (Chair)</td>
<td>Department of Health</td>
</tr>
<tr>
<td>Professor Derek Gallen</td>
<td>UK Foundation Programme Office</td>
</tr>
<tr>
<td>Professor Paul O'Neill</td>
<td>Medical Schools Council</td>
</tr>
<tr>
<td>Christine Outram</td>
<td>CEO, MEE</td>
</tr>
<tr>
<td>Dr Katie Petty-Saphon</td>
<td>Medical Schools Council</td>
</tr>
<tr>
<td>Professor Anthony Weetman</td>
<td>Medical Schools Council</td>
</tr>
</tbody>
</table>

Figure 3: Membership of the ISFP Project Board
Fifteen stakeholder organisations were represented through the ISFP Project Group, which reviewed all key products and the Project Risk Register, and provided advice and guidance for the project, on behalf of the ISFP Project Board. The membership of the ISFP Project Group is shown in Figure 4:

<table>
<thead>
<tr>
<th>Project Group Member</th>
<th>Representing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Paul O’Neill</td>
<td>Medical Schools Council</td>
</tr>
<tr>
<td>(Chair)</td>
<td></td>
</tr>
<tr>
<td>Professor Dinesh Bhugra</td>
<td>Academy of Medical Royal Colleges</td>
</tr>
<tr>
<td>Dr Ian Curran</td>
<td>Medical Educators</td>
</tr>
<tr>
<td>Dr Ashley Fraser</td>
<td>NHS Employers</td>
</tr>
<tr>
<td>Professor Derek Gallen</td>
<td>UK Foundation Programme Office and COPMeD</td>
</tr>
<tr>
<td>Martin Hart</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>Duncan Henderson</td>
<td>NHS Education for Scotland</td>
</tr>
<tr>
<td>Matthew Langdon</td>
<td>Department of Health England (on behalf of the 4</td>
</tr>
<tr>
<td></td>
<td>UK Health Departments)</td>
</tr>
<tr>
<td>Andrew Matthewman</td>
<td>Department of Health England (on behalf of the 4</td>
</tr>
<tr>
<td></td>
<td>UK Health Departments)</td>
</tr>
<tr>
<td>Professor Simon Maxwell</td>
<td>Prescribing Skills Assessment</td>
</tr>
<tr>
<td>Terry McMurray</td>
<td>NIMDTA</td>
</tr>
<tr>
<td>Professor Jim McKillop</td>
<td>Scottish Board for Academic Medicine</td>
</tr>
<tr>
<td>Dr Katie Petty Saphon</td>
<td>Medical Schools Council</td>
</tr>
<tr>
<td>Nick Deakin/ Karin Purshouse</td>
<td>British Medical Association Medical Students</td>
</tr>
<tr>
<td></td>
<td>Committee</td>
</tr>
<tr>
<td>Dr Kim Walker</td>
<td>UK Foundation Programme Office (Foundation School</td>
</tr>
<tr>
<td></td>
<td>Directors &amp; Managers)</td>
</tr>
<tr>
<td>Professor Anthony Weetman</td>
<td>Medical Schools Council</td>
</tr>
</tbody>
</table>

Figure 4: Membership of the ISFP Project Group

The project was managed by a core team in MSC: Denis Shaughnessy, Project Manager, Siobhan Fitzpatrick, Senior Policy Officer, Amy Stringer, Communications Officer, and Oliver Watson, Policy Officer. The core team produced many of the deliverables from the project, as well as coordinating the activities of the other participants and stakeholder groups, such as the individual medical schools and MSC-AA.

Funding awarded by the DH was held in a restricted account controlled by Jessica Pugh, Finance Manager of Universities UK.

5. Job Analysis of the role of the FY1 doctor

A systematic, multi-method job analysis, which incorporated a literature review, semi-structured critical incident interviews and observations, was undertaken to define the professional attributes required to be effective as an FY1 doctor. This triangulation approach was used to increase the credibility and validity of the results of the Job Analysis. A total of 294 individuals working as, or with, FY1s were involved in the Job Analysis and supporting consultation. The outcomes of this research are described in behavioural terms in the form of a Professional Attributes Framework (Figure 5), which sets out the key dimensions and example descriptors of effective performance in the role. These findings were used to determine the domains to be targeted for selection to the FY1 role, and to define the test specification for the SJT. The full report of the Job Analysis of the role of the FY1 doctor is available as Appendix D.

5.1. Methods

The literature review, which covered all existing policy documentation and relevant publications, was conducted in two parts. The first part focused on existing policy documentation and relevant publications in the literature to assist in identifying the key skills, abilities and professional attributes of an FY1 doctor. The second part of the literature review focused on SJTs and their use in high stakes selection. The results provided context for the Job Analysis of the role of
Improving Selection to the Foundation Programme

The detailed search strategy can be found in the report of the Job Analysis (Appendix D).

A total of 51 semi-structured interviews were carried out with foundation doctors, clinical/educational supervisors, Directors/Deans/Heads of Medical Education at medical schools, Trust Medical Directors, foundation school directors, Foundation Programme Directors and patient representatives. The interviews were conducted face to face (4) or by telephone (47) and lasted between 30-45 minutes. Interviewees were asked for examples of specific incidents from the FY1 job that were characteristic of effective or non-effective performance, and thus to identify the key skills and attributes needed to perform as a FY1 doctor. The interview transcripts were analysed using template analysis to identify the key themes within the data. All 51 transcripts were coded according to an initial template. The template underwent a number of iterations, and quality checks were carried out.

Observations (shadowing) of 13 FY1 doctors were carried out at four locations throughout the UK. A total of 29 hours of observations was undertaken by two researchers. The purpose of observing FY1 doctors was to gain a greater insight and understanding into the tasks and responsibilities of a FY1 doctor, as well as the professional attributes required to perform the role successfully. The researchers recorded the tasks, responsibilities, behaviours and attributes of the FY1 doctor. A record was kept of the timings of actions and behaviours to gain an understanding of the temporal nature of the role. The observations were analysed and the behaviours classified according to the template created following the interviews. Where behaviours could not be associated with existing classifications in the template, new classifications were introduced.

5.2. Professional Attribute Framework

Analysis of the data from the interviews and observations resulted in nine professional attributes and 111 behavioural descriptors being defined. These nine attributes are set out in a ‘Professional Attribute Framework’, in Figure 5, which details the behavioural descriptors that are representative of each attribute specific to the FY1 role.

<table>
<thead>
<tr>
<th>Professional Attribute</th>
<th>Professional Attribute - definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Commitment to Professionalism</td>
<td>Displays honesty, integrity and awareness of confidentiality &amp; ethical issues. Is trustworthy and reliable. Demonstrates commitment and enthusiasm for role. Willing to challenge unacceptable behaviour or behaviour that threatens patient safety, when appropriate. Takes responsibility for own actions.</td>
</tr>
<tr>
<td>2. Coping with Pressure</td>
<td>Capability to work under pressure and remain resilient. Demonstrates ability to adapt to changing circumstances and manage uncertainty. Remains calm when faced with confrontation. Develops and employs appropriate coping strategies and demonstrates judgement under pressure.</td>
</tr>
<tr>
<td>3. Effective Communication</td>
<td>Actively and clearly engages patients and colleagues in equal/open dialogue. Demonstrates active listening. Communicates verbal and written information concisely and with clarity. Adapts style of communication according to individual needs and context. Able to negotiate with colleagues &amp; patients effectively.</td>
</tr>
<tr>
<td>4. Learning and Professional Development</td>
<td>Demonstrates desire and enthusiasm for continued learning, takes responsibility for own development. Willing to learn from others and from experience. Is open and accepting of feedback. Demonstrates a desire and willingness to teach others.</td>
</tr>
<tr>
<td>5. Organisation and Planning</td>
<td>Manages and plans workload effectively, displaying efficient time management and delivering tasks on time. Able to prioritise effectively and re-prioritise where appropriate. Is conscientious and maintains accurate records.</td>
</tr>
</tbody>
</table>
### Professional Attribute - definition

<table>
<thead>
<tr>
<th>Professional Attribute</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Patient Focus</td>
<td>Ensures patient is the focus of care. Demonstrates understanding and</td>
</tr>
<tr>
<td></td>
<td>appreciation of the needs of all patients, showing respect at all times.</td>
</tr>
<tr>
<td></td>
<td>Takes time to build relationships with patients, demonstrating courtesy,</td>
</tr>
<tr>
<td></td>
<td>empathy and compassion. Works in partnership with patients about their care.</td>
</tr>
<tr>
<td>7. Problem Solving and Decision</td>
<td>Demonstrates an ability to assimilate a range of information and identify</td>
</tr>
<tr>
<td>Making</td>
<td>key issues. Engages with the wider issues and thinks creatively to solve</td>
</tr>
<tr>
<td></td>
<td>problems and reach appropriate decisions. Is proactive and demonstrates</td>
</tr>
<tr>
<td></td>
<td>initiative. Is able to attend to detail.</td>
</tr>
<tr>
<td>8. Self Awareness and Insight</td>
<td>Demonstrates awareness of the boundaries of their own competence and</td>
</tr>
<tr>
<td></td>
<td>willing to seek help when required, recognising that this is not a weakness.</td>
</tr>
<tr>
<td></td>
<td>Exhibits appropriate level of confidence and accepts challenges to own</td>
</tr>
<tr>
<td></td>
<td>knowledge.</td>
</tr>
<tr>
<td>9. Working Effectively as Part of</td>
<td>Capability &amp; willingness to work effectively in partnership with others and</td>
</tr>
<tr>
<td>a Team</td>
<td>in multi-disciplinary teams. Demonstrates a facilitative, collaborative</td>
</tr>
<tr>
<td></td>
<td>approach, respecting others' views. Offers support and advice, sharing tasks</td>
</tr>
<tr>
<td></td>
<td>appropriately. Demonstrates an understanding of own and others' roles</td>
</tr>
<tr>
<td></td>
<td>within the team and consults with others where appropriate.</td>
</tr>
</tbody>
</table>

**Figure 5: Professional Attributes Framework**

The Professional Attributes Framework was checked against the findings from the first part of the literature review. This confirmed that all of the professional attributes identified through the review had already been identified through the other two Job Analysis methods (interviews and observations), thus confirming the validity of the Professional Attributes Framework.

An online survey to validate the outcomes of the Job Analysis with a larger number of individuals gained 230 respondents. The survey was sent to a number of different organisations (with a request for it to be cascaded to their membership) and to individuals working as or with FY1s, including the BMA Junior Doctors Committee, the UKFPO Foundation Doctors Advisory Board and the UKFPO Foundation School Directors and Managers networks. The results from the validation survey indicated that all nine Professional Attributes were considered by key stakeholders to be important to be effective as an FY1 doctor and important to evaluate at selection.

The Professional Attributes Framework was mapped against existing data, including the national Person Specification and the FY1 curriculum. The mapping demonstrated good concordance between the Professional Attributes Framework and the FY1 curriculum. The mapping demonstrated that the Job Analysis and the resultant Professional Attributes Framework showed good concordance with the national Person Specification, although the Professional Attributes Framework identified one additional attribute - Learning and Professional Development - which was not specified in the FP2011 national Person Specification. Subsequently, the UKFPO Rules Group accepted the recommendation of the ISFP Project Group to include an additional statement in the FP2012 national Person Specification requiring commitment to learning and professional development (Appendix E).

The next step of the Job Analysis was to define the professional attributes that could be measured in an operational SJT, as not all can be directly measured in this format. Aspects of Organisation and Planning, and Problem Solving and Decision Making, for example prioritisation or attention to detail, are implicit in the format and preparation for an SJT and thus embedded in the test. Following the outcome of the Job Analysis, a panel (n=7) of researchers and occupational psychologists expert in SJT development and design, identified that five composite professional domains would be the most appropriate to target in an operational SJT and thus provide a psychometrically robust and valid test. Figure 6 illustrates the relationships between the nine professional attributes and the five SJT target domains.
6. SJT Item Writing and QA Process

The University of Cambridge and the Work Psychology Group were contracted to design, develop and evaluate the SJT for selection to the Foundation Programme, as described in this chapter. Full information is available in Appendix F, the report of the SJT pilots.

6.1. SJT design

The Professional Attributes Framework (Figure 5) identified through the findings of the detailed Job Analysis of the role of the FY1, and the lessons from existing use of SJTs in high stakes selection as identified through the extensive literature review, informed an initial test specification for the SJT, including the length of the test, appropriateness of content, response format and response instructions.

Ten micropilots involving (n=32) final year medical students at four medical schools (12 male, 20 female; 15 white, 17 non-white) sought feedback on the use of SJTs for selection to the Foundation Programme, the appropriateness and relevance of the content, and the test specification for response format and response instruction. The micropilots involved a short (30 minute) pilot test in exam conditions and small group discussions on 16 SJTs of four response types. All attendees were sent briefing materials in advance of the micropilot, and were asked to sign a confidentiality disclaimer on the day.

The feedback from the ten micropilot sessions was used alongside information from the literature reviews, and existing good practice from the SJT used for selection into GP training – an existing successful use of SJTs for selection into postgraduate medical training - to inform the SJT design specification used for item development and piloting. Further details can be found in Appendix F. One of the purposes of the pilots was to evaluate whether the specification was appropriate. Analysis and review of the design specification occurred throughout all stages of the development and review process.

The SJT is designed to be staged in the context of the Foundation Programme, with applicants assuming the role of an FY1 doctor in answering the items. Knowledge based instructions (‘what should you do’) are used instead of behavioural based response instructions (‘what would you do’). The evidence from the literature reviews showed that knowledge based instructions are more appropriate for high stakes selection contexts where faking and coaching are potential threats, and the GMC places a high premium on probity.
Two types of response format were used; ranking and multiple choice. The rationale for this was informed by experience in designing successful high stakes SJTs in other contexts. Using both response formats enables a fuller range of scenarios to be used, for example some scenarios require an applicant to differentiate between singular actions in response to a scenario that vary in appropriateness, whilst in other scenarios it is necessary to do more than one thing. Different response formats do not enable analysis of separate traits or skills. Whilst the ranking items allows for greater granularity in the data due to the more complex response/scoring format, they are typically more time-consuming to complete and are not appropriate for all scenarios, therefore a balance with multiple response items is needed. Approximately two thirds of items used in the test were of the ranking answer format (rank the five options in order of appropriateness), and one third were of the multiple choice format (select three from eight options). This structure also follows the validated methodology from the SJT for use in short-listing to GP training posts.

The SJT items are mapped to the five SJT target domains (Figure 6) for item development and test construction purposes, but each item does not exclusively measure a target domain in isolation. As such the SJT should not be viewed as assessing five individual domains or as a ‘blueprint’. The content of the scenario and the correct response keys are likely to relate to the targeted domain, but a number of the other response options are likely to tap into one or more of the other SJT target attribute domains. This increases the complexity of items, and avoids the possibility of a ‘template answer’ being provided by applicants.

The SJT test content should cover the range of target domains in roughly equal proportions. Similarly within and across the target domains, the test should cover a range of relevant topics e.g. dealing with a difficult colleague; being asked to undertake an unfamiliar procedure.

The SJT is designed to last for two hours and contain 60 items. In the initial pilot, 65 items were piloted in two hours; and in the large-scale pilot, 60 items were piloted in two hours. Sixty items were shown to be sufficient to cover the five target attribute domains in a sufficiently reliable way without overloading applicants. Sixty items also enables a greater distribution of scores and thus differentiation between applicants. Allowing two minutes per item is consistent with previous evidence and experience, which has demonstrated good reliability and validity.

6.2. Item development and review

In line with good practice in design and development of assessment, an individual test item should not be used until its psychometric properties are known, and the quality, relevance and validity of the question and its responses are established. The process of development and review for the SJT items is both detailed and necessary; ensuring that only those items meeting all the quality criteria are piloted. Once the results of the pilot are known, the performance of items (consistency of answers between applicants, level of difficulty etc) can be discerned. Some items may, at this stage, need further refinement and review before they can be piloted again; other items can be deemed to be the final version of the item, ready for inclusion an item bank and for use in live selection to the Foundation Programme. It is vital that the psychometric properties of live test content is known, in order to ensure fairness of assessment, and in order to facilitate standard setting or test equating between multiple papers.

There were two phases of SJT item development and review. Phase One ran between April and October 2010 and Phase Two ran between November and March 2011. Figure 7 outlines the process of the SJT item development and review process followed for both Phases One and Two. The individual stages of item development and review are described in more detail in sections 6.3 – 6.7.
6.3. Item writing workshops

Five item writing workshops were held between April and May 2010 (Stage 1 of the ISFP project) and two item writing workshops were held between November and December 2010 (Stage 3 of the ISFP project). A total of n=89 individuals from across England, Wales, Scotland and Northern Ireland were trained in SJT item-writing principles for selection to the Foundation Programme. Of these, nine were involved in item-writing during both phases of item development. The one-day item-writing workshops were accredited by the Royal College of Physicians, and attendees were awarded 6 CPD points.

Volunteer item-writers were recruited via the UKFPO Foundation School Directors, the Society for Academic Primary Care, and via a Psychiatry contact, in order to ensure that the breadth of expertise in developing items reflected a range of medical specialties (n=59 in 23 acute specialties, n=12 from 2 community specialties, 17 = not-declared) and UK regions. The national Person Specification for the role required that volunteers be familiar with the roles and responsibilities of an FY1, have worked with FY1 doctors within the previous two years, be willing to commit to attending a one day training workshop and be willing to engage in follow-up work. The demographic data of the item-writers and their job role are provided in Figure 8.

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>53</td>
</tr>
<tr>
<td>Female</td>
<td>34</td>
</tr>
<tr>
<td>Not Stated</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25 and under</td>
<td>0</td>
</tr>
<tr>
<td>26-35</td>
<td>7</td>
</tr>
<tr>
<td>36-45</td>
<td>38</td>
</tr>
<tr>
<td>46-55</td>
<td>26</td>
</tr>
<tr>
<td>56-65</td>
<td>15</td>
</tr>
<tr>
<td>66 and over</td>
<td>0</td>
</tr>
<tr>
<td>Not stated</td>
<td>3</td>
</tr>
</tbody>
</table>
All item writers were sent briefing materials in advance of the workshop, and were asked to sign a confidentiality and code of conduct form on the day. Each workshop was facilitated by a minimum of two facilitators trained in SJT item-writing principles. A member of the ISFP Project Group attended each workshop to introduce the context of the ISFP project and the role of SJTs in future selection to the Foundation Programme. The facilitators provided an overview of the test specification for selection to the Foundation Programme and provided training in SJT item-writing principles. During the workshop, item writers worked in pairs and small groups to create and review scenarios and plausible responses for the SJT items, receiving feedback from the two trainers.

In total, over the seven item writing workshops, 453 items were written, by 82 item writers. Seven item writers did not submit any items either at or following the workshops. This is a return rate of 92% and an average of 5.5 items per item writer. In the first phase of workshops, items writers were asked to aim to write between six and eight items, on the day and as follow-up work. However only two thirds of item writers actively contributed to follow-up work (for n=262 items). In the second phase of workshops, item writers were asked to write between three and five items on the day, to encourage item writers to produce better quality items and spend more time trying to refine and develop each item. In the second phase, 90% of item-writers actively contributed to follow-up work. Overall, the number of items written by each item writer ranged from one to 19. In total 265 ranking items and 198 multiple choice items were developed. This split was in line with the design specification of the SJT, which is for two thirds ranking and one third multiple choice. The number of items relevant to each of the target domains is illustrated in Figure 9.

<table>
<thead>
<tr>
<th>Target domain</th>
<th>Number of items written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment to Professionalism</td>
<td>180</td>
</tr>
<tr>
<td>Coping with Pressure</td>
<td>91</td>
</tr>
<tr>
<td>Effective Communication</td>
<td>68</td>
</tr>
<tr>
<td>Patient Focus</td>
<td>46</td>
</tr>
<tr>
<td>Working Effectively as Part of a Team</td>
<td>68</td>
</tr>
</tbody>
</table>

Figure 9: Total number of SJT items written by target domain
psychologists trained in SJT item writing principles. Where necessary, item writers were asked to provide clarification on the content of their items, and resubmit them with amendments. In some cases, there were several iterations between the item-writer and the psychologist; in most cases one or two amendments were sufficient. One of the goals of this method is to maintain buy-in from trained item writers in that they maintain ownership of the SJT items they have contributed to the programme.

Following the initial review in the first phase of workshops, 290 items were deemed to meet the quality criteria for the SJT items and were ready for the next stage of review. 115 items were deemed not suitable to be taken any further; an attrition rate of 31%. Item writers were informed of the status of the items they had written. During the second phase of workshops, 70 of the 77 items were ready for the next stage of review. Seven items were deemed not suitable and were not taken any further in the process; an attrition rate of 9%. This was a great improvement from the first set of workshops, and can be attributed at least in part to amendments made to the structure of the training.

6.5. Lead item review
The 360 items which passed the quality controls following initial review by psychologists were then reviewed by a lead item reviewer. Twelve lead item reviewers were clinicians identified during and following the SJT item writing workshops as those individuals who appeared to have fully understood the item writing principles and were willing to be involved further in the process. Whilst the SJT items written at workshops were written in pairs and reviewed in small groups, review by a separate clinician trained in SJT item-writing would ensure validity and realism to the content of the SJT item scenario and response key.

Lead item reviewers were asked to consider whether the items reflected the target domains, if they were clear and realistic, if terminology was commonly used across the UK and if the answer key was appropriate. This review stage took place electronically and remotely. Lead item reviewers had already signed a confidentiality and code of conduct form at the item-writing workshops, and they were sent no more than 25 items for review in order to minimise the security risk. Of the 360 items reviewed by lead item writers, 54 items were deemed not suitable to go any further, leaving 306 items to go forward to the next stage.

6.6. Foundation doctor focus groups
The purpose of the foundation doctor focus groups was to gain feedback from job incumbents about the relevance and fairness of the SJT items, to ensure content validity and face validity of the SJT. Twenty focus groups of between three and five foundation doctors were held in five regions of England and Northern Ireland, with a total of 63 individuals taking part. Participants were asked to sign a confidentiality and code of conduct form. The demographic data of the foundation doctors involved in focus groups are provided in Figure 10.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Not Stated</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>Mean age</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>23-39</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>White</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Mixed</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Not Stated</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 10: Focus group participant demographics
Participants were asked to review the test instructions for clarity and ease of understanding. They were invited to discuss in small groups the scenario content of individual SJT items and the response options. They were also asked to provide a possible answer key and about the timing of the test (60 items in 120 minutes). For security reasons no individual had sight of more than 20 items.

All participants in the focus groups agreed that the test instructions were clear and easy to understand. Following feedback from the focus groups, 20 of the 306 items reviewed were deemed not suitable to go forward to the next stage of review, with queries on a further 11 items. This feedback was considered alongside the concordance comments and results. In light of feedback from the foundation doctor focus groups, small revisions were made to the remaining 275 items before these were subject to concordance analysis.

6.7. Concordance Panel
Following best practice in SJT development, the aim of a concordance stage was to identify a high level of consensus between Subject Matter Experts on the answer keys to determine whether items were ready to be piloted. Where items did not achieve a high level of concordance, they were re-examined and reviewed; some were for further workshop and lead item review.

Subject Matter Experts were volunteers who met a similar person specification as for SJT item writers in that they were required to be familiar with the roles and responsibilities, skills and attributes to be successful in the role of FY1 and to have worked with FY1 doctors within the previous two years. Subject Matter Experts were recruited via the UKFPO and Medical Schools Council networks to complete a trial of SJT items in exam conditions. They were also invited to give written and verbal feedback on the content of the items. All Concordance Panel members signed a confidentiality and code of conduct form.

Four Concordance Panels of between eight and 19 Subject Matter Experts (total n=60) were conducted. Four papers containing a total of 286 items were reviewed. At this stage, the tests were not constructed as final tests (e.g. spread of domains) as the aim of the concordance panels was to analyse individual items. Using established criteria of acceptance levels, items were deemed either to have acceptable levels of concordance (259) or unacceptable levels of concordance (27). Items with above satisfactory levels of concordance were deemed suitable for piloting.

Consideration was also given to verbal and written feedback from the concordance panel when selecting items. Twenty one items (the 21 items which demonstrated satisfactory levels of concordance, rather than good or excellent) were removed from the test paper in light of feedback from the Concordance Panel. Items with lower, but still satisfactory, levels of concordance were also considered in light of the feedback from the item writer and foundation doctor focus group key. If the concordance key differed from the item writer and foundation doctor focus group key, it was not seen as appropriate for the item to be used in the pilot. Following the Concordance Panel review, 200 items were selected for inclusion in the initial and large-scale pilots.

6.8. Test construction
Test construction for the piloting phase was based on the following parameters and recommendations:

- There should be roughly two thirds ranking and one third multiple choice items within a test paper, to allow granularity of performance data and a range of scenarios
- There should be, as far as possible, an equal split between the domains being targeted
- Each item, as well as being mapped against a domain, were mapped by topic area. Scenarios were assigned to three broad topics of a ‘colleague’ topic, a ‘patient’ topic or a ‘personal’ topic. Within these three broad topic areas, scenarios were further assigned to different topic areas. A good range of topic areas should be used within each paper
- Items should be spread throughout the paper to ensure they are not grouped together by ‘topic’ or ‘domain’
- Items that include gravely serious scenarios (e.g. those involving deaths or serious mistakes by the FY1) should not be used at the beginning of the paper, as applicants may find these scenarios stressful initially; placing these items later on in the test allows applicants time to familiarise themselves with the context of the test and what is expected of them.

6.9. Scoring convention

In contrast with clinical knowledge items, with SJT items there is often no definitive correct answer. Two scoring methods were piloted: rational scoring which uses a predetermined answer key from concordance data; and empirical scoring answer key determined through use of live data. Following best practice, the SJT scoring key is based on rational scoring determined through:

- consensus at the item review stages of item writers and lead item review
- expert judgement in a Concordance Panel review
- review and analysis of the pilot data

When determining how to score each item type, consideration needs to be give to the amount of valid information which can be derived from each test item. For example, a ranking item – which takes longer for the applicant to assess and respond – should receive points for answers which are partially correct, as this provides the maximal amount of information about that applicant’s performance.

For ranking items, 20 marks were available for each item. For each of the five response options up to four marks were available. Marks were awarded for near misses. If a participant tied two options, they received 0 marks for each of the options they tied, as they have not answered the question. There was no negative marking. Although 20 marks were available for ranking items, if a participant offered a complete answer to the item, the lowest possible available score is 8 and the highest possible score is 20. The scoring key for ranking items is illustrated in Figure 11.

![Keyed Rank](https://example.com/figure11.png)  
**Figure 11: Scoring key for SJT ranking items**

For multiple response items, four marks were available for each option answered correctly, with a total of 12 marks available for each item. If an applicant answered more than three options, they scored 0 for the whole item, as it was incorrectly answered. There was no negative marking employed. This represents a lower total number of marks for multiple response items available than for ranking items, to reflect that ranking items are able to provide more detailed data than multiple response items and require more effort from participants.

Participants who did not complete an item received 0 for that item. Participants were not credited with random guesses on the items that they did not complete (identified by sporadic or unusual patterns of answers).

Scoring SJTs is a complex process and the choice of methodology for how to score the items is not straightforward. Alternatives have been considered, including not providing points for near misses, or systematically subtracting eight from each ranking item.
7. SJT Pilots

7.1. Overview
There were two phases of SJT piloting; an initial pilot in autumn 2010 and a large-scale pilot in spring 2011. The initial pilot took place in four UK medical schools, involving 455 participants. The large-scale pilot took place in 13 UK medical schools, involving 639 participants. Two non-UK pilots took place during spring 2011 involving 43 participants; however, as the sample size was so small, these data have been removed from analysis so as to avoid any potential misinterpretation. Feedback from the non-UK participants is still considered.

The purpose of the initial pilot was to ensure sufficient confidence in the psychometric characteristics of the SJT to justify a larger pilot, to inform the design and development of an SJT for selection to the Foundation Programme, and to expose any unforeseen issues with the administrative and logistical arrangements. The purpose of the large-scale pilot was to undertake further psychometric analysis to inform the recommendations as to the design and operational (live) test specification, to continue to pilot administrative and logistical arrangements, and to pilot a larger number of SJT items and therefore develop the item bank.

The pilots have demonstrated that the SJT is an appropriate method for selection to the Foundation Programme, as the pilots demonstrate construct validity and criterion related validity, and that an SJT of 60 items with the right psychometric properties is both reliable and able to differentiate sufficiently between applicants. Over 95% of participants in the pilots evaluated the SJT, and there was good consensus that it is relevant, fair and better than the current white paper questions, demonstrating stakeholder acceptability.

This chapter summarises the practical findings from the pilots and the key findings from the analysis of pilot data. Further analysis of data from the pilots is included within Appendix F.

7.2. Methods of SJT delivery
Over the two phases of piloting, the method for the delivery of the SJT was trialled on paper, online and using a hybrid method of computerised handsets, as described in detail in Figure 12. Security of SJT item content was paramount throughout the pilots.

Competitive tenders were invited for the paper-based delivery. Stephen Austin was commissioned for the printing, delivery, collection and scanning for the initial SJT pilot, and Cambridge Assessment was commissioned for the same services for the large-scale SJT pilot.

Five medical schools in the large-scale pilot volunteered to run the SJT on computers; the other schools involved in the pilots did not have the facilities to offer this to the whole final year cohort. Cambridge Assessment provided the use of, and technical support for, online software to deliver the SJT for the large-scale pilot, at zero cost to the project, as this provided an opportunity for the organisation to pilot their newly developed online delivery software.

The use of Banxia handsets, owned by the host university of one medical school involved in the initial pilot, were provided at zero cost to the ISFP project.
<table>
<thead>
<tr>
<th>Description</th>
<th>Paper</th>
<th>Computer</th>
<th>Handsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 1 x 36pp question booklet</td>
<td>• Secure login with timed access (centrally controlled)</td>
<td>• Battery operated individual handsets were used, and remotely connected to a ‘hub’</td>
<td></td>
</tr>
<tr>
<td>• 1 x 2pp OMR answersheet</td>
<td>• Answers were recorded in real-time</td>
<td>• Participants completed the paper-based pilot and then entered their answers onto the handsets (extra time to do so)</td>
<td></td>
</tr>
<tr>
<td>• 1 x 2pp glossary</td>
<td>• Cambridge Assessment staff attended all pilot sites to provide technical support</td>
<td>• Answers were entered in real-time and then uploaded in one go</td>
<td></td>
</tr>
<tr>
<td>• Applicants completed the OMR answersheet; the papers were then scanned</td>
<td>• Papers were provided as a back-up, but were not used</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and answers converted to a .csv file</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>participants</td>
<td>• 455 participants in the initial pilot from four UK schools; 421</td>
<td>• 218 participants from five UK schools in the large-scale pilot completed the SJT on computers</td>
<td></td>
</tr>
<tr>
<td>participants</td>
<td>participants from eight UK schools and 43 non-UK participants from</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>two non-UK schools in the large-scale pilot completed the SJT on</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applicants with special</td>
<td>• Coloured papers and enlarged papers were provided by request</td>
<td>• No adaptations were made. The screen size is fixed, and it would be difficult to adapt. The only feasible option would be for applicants with special needs to record their answers in another format, for these to then be manually entered</td>
<td></td>
</tr>
<tr>
<td>requirements</td>
<td>• Extra time was arranged and facilitated by the pilot sites</td>
<td>• Separate hubs could be used for sessions of different test lengths</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>• Papers were securely stored in tamper-evident bags by the printer.</td>
<td>• Secure network and secure storage of content (items) assured throughout</td>
<td></td>
</tr>
<tr>
<td>(summary)</td>
<td>One working day in advance of the pilot, boxes of papers were</td>
<td>• It was not possible to take a screen shot, press ‘escape’ or the back button</td>
<td></td>
</tr>
<tr>
<td></td>
<td>couriered to a named contact at the medical school, and signed for.</td>
<td>• The administrator had to approve each participant to join the session (only recognised users)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Boxes of papers were stored overnight according to local medical</td>
<td>• Unique login details were provided on desks – to prevent anyone accessing the server from outside of the room</td>
<td></td>
</tr>
<tr>
<td></td>
<td>school protocol for secure storage. Papers were counted out and in</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>to ensure that all copies of all papers were returned</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Administration (preparation)
- Arranging secure overnight storage of boxes of papers

### Administration (on the day)
- Papers were provided on the desks before participants entered the room
- Example items provided
- Computers were logged in to the secure server when participants entered the room
- Short briefing on how to use the software, plus opportunity to practise use of software
- At one school the computer server was re-booted as the test was about to start; all participants were re-logged back onto the system within 15 minutes
- Incompatibility with screen size at one school – participants had to scroll

### Venues
- Typical exam halls were used – desks sufficiently spaced to allow invigilators to walk between desks
- In some venues, computers were widely spaced, or participants sat at alternate desks. All desks faced the same direction
- Difficult to administer extra time – whilst access could be provided to applicants within the same session, they needed to end the session together to minimise disruption

### Participant feedback
- There was no negative participant feedback
- Some participants commented that they were not aware the glossary was available (it was indicated in the briefing and included as a separate paper)
- UK participants were familiar with the format of the OMR forms and there were rarely comments about the mode of delivery; not all
- Overall, participants found the software easy to use, navigate, clear where the glossary was, and clear which answers had been attempted and/ or completed
- There was a mixed reaction to the ease of use of the drag & drop answers
- Most negative feedback related to the pilot site which had experienced the server
- Overall, quite negative feedback
- Applicants were unfamiliar with the use of handsets, and it took some time to get used to entering the data
- Participants commented that it was difficult to scroll between answers, and there was no indication if an answer had been completed or not

### Paper
- Arranging secure overnight storage of boxes of papers

### Computer
- Arranging secure overnight storage of boxes of papers
- The IT lead for the pilot site spent circa 2 days' time pre-loading computers with access to the software via the network.
- The length of test and question formats had to be pre-loaded onto a ‘session’ – this took circa 2 days’ admin time

### Handsets
- Arranging secure overnight storage of boxes of papers

- Handsets were distributed after questions papers were collected, but with answer sheets still on desks

- As for paper-based pilots
<table>
<thead>
<tr>
<th>Pros</th>
<th>Computer</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>auto-generated answers</td>
<td>participant and invigilator unfamiliarity</td>
</tr>
<tr>
<td></td>
<td>no issue about being unable to read</td>
<td>difficulties in navigating through the questions – other handsets are available</td>
</tr>
<tr>
<td></td>
<td>handwriting for RA numbers etc</td>
<td>recommendation that applicants completed both paper-based and entered into handsets is both time-consuming and additional resource</td>
</tr>
<tr>
<td></td>
<td>prompts for incorrect number of answers given or if partial answer</td>
<td>issues of providing individual adaptations (extra time etc)</td>
</tr>
<tr>
<td></td>
<td>auto-generated answers</td>
<td>papers would need to be printed and provided to schools as a back-up – this would add to cost</td>
</tr>
<tr>
<td>participation familiarity with OMR sheets</td>
<td></td>
<td>participant and invigilator unfamiliarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>difficulties in navigating through the questions – other handsets are available</td>
</tr>
<tr>
<td>All UK schools routinely run paper-based assessments for large numbers; with sufficient notice, there is no difficulty in arranging for a large, flat, well-lit hall with sufficient capacity</td>
<td>Fewer than one third schools have facility to deliver assessments via computers for whole cohort in only one or two sittings – even when allowing for booking several computer rooms for use in parallel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>issues of compatibility of system with multiple operating systems and screen sizes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>issues of providing individual adaptations (extra time etc)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>papers would need to be printed and provided to schools as a back-up – this would add to cost</td>
</tr>
<tr>
<td>non-UK participants were familiar with the concept of the OMR form, however the sample questions helped with this</td>
<td>reboot</td>
<td>participant and invigilator unfamiliarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>difficulties in navigating through the questions – other handsets are available</td>
</tr>
<tr>
<td>A small number of participants suggested that the SJT might work well on computers</td>
<td>It should be noted that the participants of the computer tests were those participants who already routinely use computers for medical school assessments</td>
<td>recommendation that applicants completed both paper-based and entered into handsets is both time-consuming and additional resource</td>
</tr>
<tr>
<td></td>
<td></td>
<td>issues of providing individual adaptations (extra time etc)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>papers would need to be printed and provided to schools as a back-up – this would add to cost</td>
</tr>
</tbody>
</table>

Analysis of the mean scores for participants taking the SJT as paper-based and computer-based showed no statistical difference, indicating that the mode of delivery did not affect the answers. Analysis of the raw answers entered by handsets and the answers entered by the same participants on paper did not identify any differences – however it would be difficult to defend which set of answers should count in the event of a discrepancy.

![Figure 12: Methods of SJT test delivery](image)
7.3. Administration

Guidance for invigilators was developed by the ISFP project team, informed by existing protocols for universities and national examination bodies, to support standardised administration at the pilot sites. The guidance document included recommendations around the number of invigilators, the venue to be used, the applicant briefing to be read to participants, as well as policies to address e.g. a security breach. The Administrators guidance is provided as Appendix G.

A tailored administration document which included a background to the project, instructions for administration on the day, invigilator checklists, a briefing for participants and a feedback sheet was sent to all named leads for the pilot sites. Tailored versions of these materials were developed for those schools delivering the SJT on paper, computer and handsets, and for the one school completing a NEO PI-R personality measure. A reference guide was also produced by Cambridge Assessment for administrators who would be using the online software.

Staff and students involved in the pilots were asked to treat the SJT pilots as a live selection test, under examination conditions. The majority of participants in the initial and large-scale pilots indicated on their evaluation forms that the SJT was well run and well invigilated. Feedback from SJT administrators informed the running of subsequent pilots. There were a small number of administrative issues, the lessons of which are summarised in Figure 13.
<table>
<thead>
<tr>
<th>Pilot findings</th>
<th>Lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venues</td>
<td>SJTs must be taken in exam conditions in a typical exam venue, with sufficient space to walk between desks. Venues to provide details in advance</td>
</tr>
<tr>
<td>Initial pilot: Two pilots took place in lecture theatres – this made it impossible for invigilators to walk between desks, and for students to leave to go to the bathroom without disturbing others. In at least one venue there was no clock in the venue</td>
<td>Clocks must be functioning and visible</td>
</tr>
<tr>
<td>Large-scale pilot: Participating schools were informed of the detailed requirements in advance. Most venues were suitable. However, the computer room at one venue had high background noise and two venues had no clock visible</td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td>Nominated responsible officer at each pilot site</td>
</tr>
<tr>
<td>Initial pilot: At one pilot there appeared to be no lead invigilator – the MSC ‘observer’ assumed this role</td>
<td>Nominate academic ‘champion’ for SJT</td>
</tr>
<tr>
<td>Large-scale pilot: Lead invigilators were in place at each participating school. The involvement of academics was beneficial for student engagement</td>
<td></td>
</tr>
<tr>
<td>Student behaviour and invigilation</td>
<td>Invigilators at several sites were briefed by the lead administrator prior to the pilot.</td>
</tr>
<tr>
<td>Initial pilot: At some sites a small number of students did not treat the pilot as seriously as a real application – chatting when finding a seat, leaving the test early. At others, students were allowed to leave the test venue early before the end of the session. In at least one case, the instructions for applicants were not read aloud at the start of the test</td>
<td>Briefing to students should be revised to make sure that it is an appropriate length</td>
</tr>
<tr>
<td>Large-scale pilot: Student behaviour was good at most sites though a small number were talking as they took their seats at some pilot sites. Invigilators at many sites reported that the briefing to applicants was too long; this led to some invigilators abbreviating the information</td>
<td></td>
</tr>
<tr>
<td>Instruction - ‘should’ or ‘would’</td>
<td>Ensure the briefing and written instructions are clear about the significance of the ‘should’ instruction</td>
</tr>
<tr>
<td>Initial pilot: A number of students were confused around the importance of the ‘should’ instructions</td>
<td></td>
</tr>
<tr>
<td>Large-scale pilot: Adjustments to the applicant briefing and test materials appear to have been successful</td>
<td></td>
</tr>
<tr>
<td>Concern about regional variance</td>
<td>Ensure that concordance and review of items is undertaken by individuals in a range of specialties and a range of geographic</td>
</tr>
<tr>
<td>Pilot findings</td>
<td>Lesson</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>clinical settings and between different regions</td>
<td></td>
</tr>
<tr>
<td>Large-scale pilot: This concern did not arise in feedback sessions</td>
<td>regions</td>
</tr>
<tr>
<td><strong>Access arrangements</strong></td>
<td></td>
</tr>
<tr>
<td>Initial pilot: No requests to adjust for access arrangements apart from enlarged paper size at one pilot site</td>
<td>Establish special requirements of applicants at an early stage and build in to delivery specifications</td>
</tr>
<tr>
<td>Large-scale pilot: A variety of adjustments were required including extra time, different coloured paper, coloured overlays, enlarged paper, enlarged font sizes (min size 14pt), different font types and rest breaks</td>
<td>Financial implications of the provision for multiple access arrangements to be considered</td>
</tr>
<tr>
<td><strong>RA numbers</strong></td>
<td></td>
</tr>
<tr>
<td>Initial pilot: Some students mis-entered or left blank their RA numbers</td>
<td>Ensure that candidates understand anonymity is ensured. Invigilators to check papers against master list of RA numbers</td>
</tr>
<tr>
<td>Large-scale pilot: There was some mis-entering and missing RA numbers, however all students made aware of anonymity</td>
<td>Reinforce importance of accuracy to applicants</td>
</tr>
<tr>
<td><strong>Guidance for administrators</strong></td>
<td></td>
</tr>
<tr>
<td>Initial pilot: Some administrators wanted more modular guidance based on different roles on the day</td>
<td>Have separate checklists for i) lead invigilator and ii) other invigilators. Seek to shorten the guidance for administrators where possible</td>
</tr>
<tr>
<td>Large-scale pilot: Some felt that documentation to be overly detailed</td>
<td></td>
</tr>
<tr>
<td><strong>Test paper delivery (courier)</strong></td>
<td></td>
</tr>
<tr>
<td>Initial pilot: The test papers were not delivered by the couriers to responsible officer; instead they were handed over to some other person at the medical school</td>
<td>Ensure that the contractual arrangements with the couriers contain clear instructions for delivery</td>
</tr>
<tr>
<td>Large-scale pilot: All papers delivered to responsible officer</td>
<td></td>
</tr>
<tr>
<td><strong>Computer delivery of the SJT</strong></td>
<td></td>
</tr>
<tr>
<td>Large-scale pilot: At one school the computer server was rebooted as the test was about to start (all participants were logged back onto the system within 15 minutes). Incompatibility with screen size at one school – participants had to scroll</td>
<td>Ensure technical support is on-hand</td>
</tr>
<tr>
<td></td>
<td>Continue to provide papers for SJT use as a back-up</td>
</tr>
<tr>
<td><strong>Scoring key</strong></td>
<td></td>
</tr>
<tr>
<td>Initial pilot: Students wanted more of a guide around how marks are allocated and how to split their time effectively</td>
<td>Share key elements of marking scheme</td>
</tr>
<tr>
<td>Large-scale pilot: Scoring key shared with students in feedback</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13: Administrative lessons for SJT delivery
7.4. Participation
Participation in the SJT pilots was voluntary. Responses and feedback was anonymous; individual feedback on performance was not provided, as the test was still in development. A minimum participation of 1,000 students was sought in order to provide statistical confidence in the findings of the SJT pilots. A minimum sample size for the psychometric analyses for each paper was n=100, which was achieved for every SJT paper, as illustrated in Figure 14 overleaf. Given this, there can be a reasonable degree of confidence in the findings of the pilots.

The initial pilot in autumn 2010 highlighted a lower than hoped participation amongst students. The timing of the pilots was the most frequent explanation for this, with many participants being about to take or having just taken final exams – or away on a ‘reflection week’. Additional schools were asked to run pilots, and other measures were taken to maximise student awareness of the role of the pilot SJTs, and incentives provided to ensure participation. Workshops (briefing sessions) on the ISFP pilots were delivered in ten of the fifteen UK medical schools, and this noticeably increased participation in the pilots. Other incentives to participate in the pilots included:

- The experience of taking part in an SJT as preparation for future selection processes
- Feedback on the SJT and answers to a subset of items
- A random cash prize draw (one per school)
- £1,000 provided to student organisations for 50% turnout (large-scale pilot only)

7.5. Feedback from pilots
Feedback was sought from participants involved in the pilots through an evaluation questionnaire. A member of the ISFP project team or Work Psychology Group attended the majority of SJT pilot sessions to provide a feedback session to participants on the SJT and answers to a subset of questions, using a standardised slide-set. The feedback session was an opportunity to explain the role of the SJT pilot in developing future selection methods, to dispel any concerns raised, and to hear the views of participants. Overall feedback was positive, particularly when comparing the SJT with ‘white space’ questions.

During the initial pilot, evaluation forms were completed by 428 participants. 78.7% of participants agreed or strongly agreed that the content of the test seemed relevant to the FP. 67.7% agreed or strongly agreed that the content of the test appeared to be fair. Only 32.1% of participants agreed or strongly agreed that the results of the test should help selectors to differentiate between weaker and stronger participants, however 31.5% neither agreed nor disagreed with this statement.

During the large-scale pilot, evaluation forms were completed by 652 participants; this includes some participants from the non-UK sample who took the initial pilot paper. 89% of the participants felt that the test instructions were clear and easy to understand, while 85% thought that the information given about the pilot was clear and helpful. 85% of participants also agreed or strongly agreed that the content of the SJT seemed relevant to the FP. 80% felt that the scenario content was appropriate for their level of training and 75% considered that the difficulty level was appropriate. Overall, 73.3% of participants agreed or strongly agreed that the content of the test was fair, but when considering whether the results of the test would help differentiate between the strong and weak participants, 42% neither agreed nor disagreed.

The majority of participants commented that the SJT is a better method of selection than ‘white space’ questions. Only a very small number of participants felt that the SJT should not be adopted in its current form, and these participants mostly believed that sufficient improvements could be made to make the SJT a useful selection tool. Feedback suggested that participants felt the test was about the right length in terms of time and content, though they often felt that there was a large amount of information to read in the time. It has been suggested that the stage that students are at in their course could affect the answers they give. Several feedback sessions have raised a lack of awareness amongst participants about the different responsibilities of clinical supervisors and educational supervisors.
7.6. Initial pilot - key findings

Figure 14 outlines the gender, ethnicity and nationality of the 455 participants at four UK schools who took part in the initial pilot. The mean age of the sample was 24.7 years, with a range of 22-43 years.

<table>
<thead>
<tr>
<th></th>
<th>No. of participants</th>
<th>% of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>159</td>
<td>34.9%</td>
</tr>
<tr>
<td>Female</td>
<td>275</td>
<td>60.4%</td>
</tr>
<tr>
<td>Not declared</td>
<td>21</td>
<td>4.6%</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>319</td>
<td>70.1%</td>
</tr>
<tr>
<td>Asian</td>
<td>67</td>
<td>14.8%</td>
</tr>
<tr>
<td>Black</td>
<td>10</td>
<td>2.2%</td>
</tr>
<tr>
<td>Chinese</td>
<td>10</td>
<td>2.2%</td>
</tr>
<tr>
<td>Mixed</td>
<td>12</td>
<td>2.6%</td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
<td>1.5%</td>
</tr>
<tr>
<td>Not declared</td>
<td>30</td>
<td>6.6%</td>
</tr>
<tr>
<td><strong>Nationality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>British</td>
<td>378</td>
<td>83.1%</td>
</tr>
<tr>
<td>Non-British</td>
<td>30</td>
<td>6.6%</td>
</tr>
<tr>
<td>Not declared</td>
<td>47</td>
<td>10.3%</td>
</tr>
</tbody>
</table>

Figure 14: SJT pilot participant demographics (initial pilot, Autumn 2010)

The pilot paper consisted of 65 items (42 ranking and 23 multiple choice) to be completed in two hours. Five of the items were used to give feedback to participants. The total available score for the paper was 1116. Analysis was adjusted to exclude erratic answer patterns and large amounts of missing data.

The majority of participants (94.9%) completed the test within two hours, supporting the intention that the SJT is a ‘power test’ not a ‘speed test’. However it should be noted that attitude to a pilot which does not count towards their live application is likely to differ from a live application process, where applicants could be expected to spend longer reflecting and reviewing their answers. Statistical analysis demonstrated a high level of internal reliability ($\alpha=0.74 – 0.77$), only slightly below the level of internal reliability required for a high stakes assessment ($\alpha=0.80$). Psychometrically, over half of the items worked well. The items that did not perform well psychometrically are undergoing further review and refinement.

Two scoring conventions were considered – rational scoring, whereby the scores are determined using a pre-determined answer key, and empirical scoring, where items are administered to a large-scale pilot sample. For items where the empirical key differs from the rational key, agreement was sought with Subject Matter Experts as to whether the empirical key is appropriate. There was a wide range of scores in the initial pilot (minimum 682 and maximum 952 for rational scoring; minimum 710 and maximum of 992 for empirical scoring). This range of scores demonstrates that although there were a number of tied scores, overall the test is able to differentiate between applicants.

Female participants scored slightly higher than male participants (by approximately 0.3 standard deviations). White participants scored slightly higher than Black and Minority Ethnic (BME) participants. This is a typical and comparable finding with many selection methods and examinations. Caution should be employed when interpreting these results due to the small sample sizes.

All participants in the initial pilot were required to complete a work style evaluation questionnaire, which provides insight into social desirability. Work using the Social Desirability scale suggests that applicants ‘faking’ their responses to reflect their perception of desirable rather than what they would do in a situation was not a serious issue. In theory
applicants cannot prepare for the test, although there is an inevitable risk that private companies set up ‘revision’ courses and guides.

The findings from the initial pilot indicated that the majority of items worked well psychometrically, the reliability of the test was good and it was able to differentiate between participants. The pilots confirmed the design specification for the SJT, and provided lessons for the administrative and logistical delivery of the SJT.

7.7. Large-scale pilot - key findings

Figure 15 outlines the gender, ethnicity and nationality for the 639 participants in the large-scale pilot. The mean age of the sample was 24.8 years, with a range of 21-48 years. Fourteen participants (2.2%) self declared that they had a disability, as defined by the Equality Act 2010.

<table>
<thead>
<tr>
<th>No. of participants</th>
<th>% of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>228</td>
</tr>
<tr>
<td>Female</td>
<td>401</td>
</tr>
<tr>
<td>Not declared</td>
<td>10</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>436</td>
</tr>
<tr>
<td>Asian</td>
<td>89</td>
</tr>
<tr>
<td>Black</td>
<td>14</td>
</tr>
<tr>
<td>Chinese</td>
<td>31</td>
</tr>
<tr>
<td>Mixed</td>
<td>23</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
</tr>
<tr>
<td>Not declared</td>
<td>22</td>
</tr>
<tr>
<td><strong>Nationality</strong></td>
<td></td>
</tr>
<tr>
<td>British</td>
<td>553</td>
</tr>
<tr>
<td>Non-British</td>
<td>69</td>
</tr>
<tr>
<td>Not declared</td>
<td>17</td>
</tr>
</tbody>
</table>

Figure 15: SJT pilot participant demographics (large-scale pilot, Spring 2011)

The large-scale pilot paper consisted of three papers, in order to maximise the number of new SJT items piloted. Each paper consisted of 60 items (40 ranking and 20 multiple choice) to be completed within two hours. Within each of the papers, 15 of the items were anchor items common to each paper to assist with analysis, plus the same five items used in the initial pilot for feedback. At this stage, it was not possible to use these to equate the tests as this was the first time the items had been piloted. A total of 1,040 marks were available for each version of the SJT paper.

As outlined in Figure 16, there was an unequal split of participants between the three papers. Efforts were made to ensure an even number of participants across the three papers, although the number of participants at each pilot site could not be predicted.

<table>
<thead>
<tr>
<th>No. of participants</th>
<th>% of overall sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paper One</strong></td>
<td>138</td>
</tr>
<tr>
<td><strong>Paper Two</strong></td>
<td>344</td>
</tr>
<tr>
<td><strong>Paper Three</strong></td>
<td>157</td>
</tr>
</tbody>
</table>

Figure 16: Participation in three papers of large-scale pilot

In total, 614 participants (96.0%) completed all items, confirming that the SJT is a power test, and that two hours is an appropriate amount of time to complete 60 SJT items.
Analysis was adjusted to remove participants who had an erratic answer pattern or a large amount of missing data. For all three papers, test statistics showed a relatively high level of internal reliability ($\alpha=0.71 - 0.76$), before being adjusted to exclude poorly performing items. When only those items that have good psychometric properties were analysed and corrected for test length using the Spearman Brown formula, all three papers show good internal reliability ($\alpha=0.79 - 0.85$), sufficient for that required for a SJT being used in a high stakes assessment ($\alpha=.80$).

The range of scores (minimum 658 and maximum 922) is similar to that achieved in the initial pilot. This is a good spread and shows that the test is able to differentiate between participants. Analysis showed a high level of agreement between the rational keying and empirical keying ($r=.96$, $p<.001$).

Female participants scored slightly higher than male participants (by approximately 0.5 standard deviations) on Paper One; there were no significant differences by gender on Paper Two and Three. White participants scored higher than BME participants on all three tests. This is a typical and comparable finding with many selection methods and examinations. It should be noted that there was variation between the papers as to how gender and ethnicity affected results, suggesting that a strong pattern of group differences cannot be identified. Further research could help to explain observed group differences.

Around 100 of the 135 items in the large-scale pilot were of sufficient quality to be used in future tests. The remaining items require further review and refinement, as they did not achieve the psychometric properties for a live SJT item in the large-scale SJT pilot. The number of items that were of sufficient quality for future use is typical for SJT development and further review will help to identify the reasons for poorly performing items.

In order to evaluate the construct and criterion related validity, participant performance (tested on two small separate cohorts) on a NEO PI-R personality measure, and comparison with OSCE performance and written assessments, were evaluated. The pilots demonstrated positive correlation with high levels of extraversion, openness to values and achievement. From this, it can be concluded that the SJT broadly measures attributes relevant to the skills needed for work in the Foundation programme.

8. Educational Performance Measure (EPM)

Following extensive consultation and piloting, all stakeholders have reached agreement on a standardised EPM framework of performance at medical school in relation to the graduating cohort up to the point of application to the Foundation Programme, which meets the agreed quality criteria. This next chapter details the process of consultation and piloting, the lessons learned, and decisions taken in developing the EPM framework. The final EPM framework is provided in Appendix I.

8.1. Background

Applicants to the Foundation Programme currently receive a score based on performance at medical school in relation to their cohort, as ranked into four quartiles worth 40, 38, 36 and 34 points. This score is then combined with their score from an online application form, ‘white space’ questions (total of 60 points), which includes 10 points for degrees, presentations, prizes and publications.

Concerns about the use of academic quartiles, raised before and during the detailed Option Appraisal, include the comparability of applicants from different medical schools and discrimination of applicants at the margins between quartiles. Given issues around the lack of transparency and lack of consistency in the allocation of points for quartiles across medical schools, one of the main drivers to review the quartile system was to ensure defensibility in the event of legal challenge. Stakeholder feedback showed strong support for the use of some measure of academic performance as well as non-academic and possibly extra-curricular activities in the selection process. The International Expert Panel, whose advice was sought during the 2009 Options Appraisal, supported the principle of
making greater use of information accumulated during medical school, to recognise that the Foundation Programme is an extension of education and training, both to value applicants' academic achievements, and using the evidence of past performance as the greatest predictor of future performance.

The ISFP Project Group (previously the Steering Group) recommended that a standardised measure of applicant educational performance up to the point of application to the Foundation Programme be piloted and developed. It was envisaged that the EPM would be derived using a standardised and transparent framework of existing performance measures, agreed with medical schools in consultation with applicants and other stakeholders, and that this measure would address some of the concerns around comparability of applicants from different schools, and that it would enable greater granularity when combined with the SJT. All UK and non-UK medical schools would be required to provide an EPM score to the UKFPO.

Under the current system of applications to the Foundation Programme, applicants receive up to 10 points for previous degrees, presentations, prizes and publications which meet the criteria specified by the UKFPO. These points are awarded within the 'white space' questions. There was full stakeholder support for recognising the most exceptional applicants and their additional educational achievements as evidenced by degrees, presentations, prizes and publications, and incorporating this into the EPM.

8.2. Consultation and piloting

There were two in-depth consultations with all 31 UK medical schools to identify the type of information on student performance available at the point of application to the Foundation Programme, and the practicalities of using this within the EPM framework. At this stage, wider stakeholder consultation was not sought.

The first consultation, in September and October 2009, gathered evidence around the information currently used to inform quartile rankings, and the assessment information on applicant performance currently collected and utilised by UK medical schools. A summary of responses is provided in Appendix H, and can be summarised as follows:

- All schools record data on student performance relating to clinical skills and curriculum knowledge. Measures of professionalism and records of extra-curricular activities are rarely collected systematically, and are impossible to quantify.
- There is a range of assessments covering clinical skills, curriculum knowledge, Student Selected Components (SSCs) and measures of professionalism, as well as summative, formative and progress-testing, used by medical schools. The timing at which the subject knowledge is assessed varies between schools, as different curricula approach the Tomorrow's Doctors outcomes framework in different ways.
- Assessment formats include Single Best Answer (SBA), Multiple Choice Questions (MCQs), Extended Matching Questions (EMQs), Observed Structured Clinical Examinations (OSCEs), case presentations and essays. The number of assessments of each type, and the weightings between the different curriculum elements, varies considerably between medical schools.
- The GMC framework Tomorrow’s Doctors specifies the minimum requirements of a competent doctor, and as such, the majority of assessments devised by medical schools are designed to assess whether a student’s performance is satisfactory or not, rather than to differentiate between the most exceptional students and those closer to the borderline. Given this, a full ranking within a cohort is neither desirable nor feasible. Some assessments, for example Student Selected Components, are typically pass/ fail. For other assessment types, a grade is awarded - often in line with university policy -, and for others, a raw score or percentage.
- As part of the undergraduate MBChB or MBBS programme, students have the opportunity to undertake additional study towards an intercalated degree, usually taking an extra year of study to complete. Intercalation is usually by competitive entry, based on performance at medical school, and is not available to all students. Between 0% (Warwick – all graduate entry) and 100% (Cambridge, Imperial, Oxford – excluding graduate entrants) of students at a given school intercalate), with a median of 25% across the country.
Analysis of the initial consultation led to the proposal that in place of overall quartiles to reflect performance in relation to the cohort, a more fair and granular method would be to record two quartile measures of performance – one relating to clinical skills or practical assessments, and a second relating to the underpinning scientific knowledge or written exams. The two scores would then be combined to provide an overall score, with additional granularity to the current quartile points.

A second consultation, in November and December 2009, sought educationalist opinion around the principles of an EPM framework on the basis of two quartile measures, including the relative weighting between the two components, the weighting between the different years of study, as well as more detailed questions around how additional points for additional academic achievements (degrees, presentations, prizes and publications) might be awarded. The responses are summarised as follows:

- Given that medical schools have diverse approaches to delivering the undergraduate curriculum and a range of assessments of student performance on the required outcomes, it should be the responsibility of individual schools to determine how many, which, and the relative weightings of, assessments to be representative of educational performance at the point of application to the Foundation Programme.
- The weighting between measures of clinical ‘practical’ skills and ‘written’ curriculum knowledge should be 50:50. Some schools did raise concerns about the practicalities of providing separate scores for these two elements.
- Performance across all years of study should be recognised, with a greater emphasis on performance in the later years of study. Differences in the length of course – for example between fast-track graduate entry (4 years) and standard entry (5/6 years) - should be taken into account.
- The majority (25) of schools believed that the EPM should recognise additional academic achievements (presentations, prizes, and publications) in addition to performance in relation to the cohort. There were also proposals to recognise the most exceptional students within the cohort.
- There was mixed opinion around whether the EPM should award points for possession of degrees in addition to the MBChB or MBBS, with 13 schools (7 schools with Graduate Entry (GE)) commending additional points for previous degrees; 7 (4 GE) schools commending points only for degrees gained during undergraduate medical degree; and 6 (3 GE) schools commending no points for additional degrees.

8.3. Pilot of draft EPM framework

A draft EPM framework for piloting, informed by the two consultations, was agreed by the ISFP Project Group. The pilot EPM framework called for applicants to be given two scores - one relating to written assessments and the other relating to practical assessments – according to a specific prescription of weightings between earlier (40%) and later years (60%) of the course. The pilot EPM framework also specified rules that:

- Only data from summative assessments, not formative ones, would count.
- Where re-sits have been taken, the marks would be capped at the pass mark except in the case of extenuating circumstances.
- Schools would devise appropriate weightings for modules and assessments.
- Schools whose students transfer to another school partway through the course would be required to provide the school of graduation with the necessary data to allow early years’ data to be included.
- The calculations would be made using standard-set raw marks, not grade points.

Of the 25 UK medical schools involved in the EPM pilot in May 2010, 22 schools provided performance data for 5,373 applicants to FP2010 in the form of separate scores for written and practical assessments. EPM data were provided as normalised scores, to allow comparisons to be made within a single school cohort. Three schools provided EPM scores for a further 442 students in different cohorts, although the scores could not be correlated with other information. The remaining six medical schools indicated that they would in principle be able to adhere to the draft
framework, but had not been able to do so within the timeframe. As the pilot used retrospective data, the EPM scores could then be analysed with the original quartile scores supplied by the medical school, and with the application form scores achieved by applicants to the FP2010. The report of the pilot of the draft EPM framework is provided in Appendix H, and the findings can be summarised as follows:

- For some schools, the pilot EPM framework required the use of additional assessment data (for example during the early years of study) not usually included in the calculation of quartiles; other schools reported that the framework limited the number of assessments that could be used (for example by requiring a split between ‘written’ and ‘practical’ EPM scores). Typically, between three and fourteen assessment scores were used across the two measures of performance; some of these were raw scores but others were on the basis of grade points.

- To illustrate the diversity of scores provided by participating medical schools:
  - One school provided written EPM scores consisting of 13 possible scores, for 228 applicants, between 0.5 and 2 (2 being the lowest performing applicant and 0.5 the best performing applicant) whereas another school provided written EPM scores as 210 possible scores (223 applicants) within a range of 139.6 and 397.6.
  - Eight schools were able to provide a full discrete ranking for the cohort, although there was feedback that this level of differentiation was artificial.
  - Twelve schools provided written EPM scores that enabled division into four equal sized quartiles. Owing to tied scores, with one exception, for the remaining schools there was a difference in group size of between four and ten applicants.
  - Three schools provided practical EPM scores which enabled a full rank (no tied scores). For 16 of 25 schools, applicants could be ranked into four equal quartiles. Owing to tied scores, with one exception, for the remaining schools there was a difference in group size of between four and ten applicants.

- The Pearson’s correlation co-efficient was calculated to correlate the written EPM scores and practical EPM scores with applicants’ original quartile points score. Across 22 schools, the written EPM scores had a strong correlation of 0.86, but across 20 schools there was only a moderately strong correlation of 0.65 for the practical EPM scores with the original quartile score. This indicates that the EPM pilot framework places a different emphasis on the elements currently used to inform quartile points.

- The Pearson’s correlation co-efficient for written EPM scores and practical EPM scores was 0.51, with just 2,306 of the 5,373 applicants placed in the same quartile for both measures, and 899 applicants (17%) ranked more than two or more quartiles differently (e.g. 1st in one measure and 3rd or 4th in the other).

- For comparison of quartile scores with written EPM scores 67% of students were in the same quartile; 30% of students moved to one quartile either side of their original rank; and 3% of students moved by more than 2 quartile ranks. For comparison of quartile ranks with practical EPM scores 46% of students were in the same quartile; 44% of students moved to one quartile either side of their original rank; and 10% of students moved by more than 2 quartile ranks.

Different methods of combining scores for the written EPM scores and practical EPM scores were modelled, including addition, multiplication, division and standard deviation. Whilst distance from the mean is the statistically more robust measure, the unknown variables (difficulty of assessments, comparability between schools) cast doubt on the reliability of any measure other than combining two scores by addition.

By adding together the points for written EPM quartiles and practical EPM quartiles, the range of points achieved ranged from four possible overall bandings to seven possible bandings. Further analysis of the underlying data indicated that it would be possible to further increase the differentiation between applicants (granularity) by extending the different number of bandings of points available for performance at medical school from four to seven, and again from seven to ten. Increasing the number of bandings can help to differentiate applicants by i) providing a greater...
spread of scores and ii) reducing the number of applicants with tied scores, without greatly affecting the values of scores achieved in terms of mean, median or mode.

8.4. EPM pilot lessons
The lessons of the EPM pilot can be summarised as follows:

- The EPM pilot demonstrated the importance and the benefits of a nationally agreed framework in terms of transparency and consistency
- Averaging an applicant’s performance over a ‘basket’ of assessments should be more representative of the applicant’s capability than taking any one of the assessments individually
- The EPM pilot highlighted that a framework which specifies the proportion of assessments from different stages in medical school is inappropriate, since it does not recognise the variation across schools in the timing of courses and assessments. It is not practical to specify the exact composition of the EPM, and the different approaches to assessment and learning should be encouraged
- An EPM framework that calls for a clear split between assessments of curriculum knowledge and clinical skills is not feasible, as one competency underpins the other
- Not all schools were able to adhere to the EPM pilot framework in full, as the university systems did not record raw scores only grades in some instances, and it was not possible to identify re-sit marks for three or four years previously. An EPM framework would need to be sufficiently flexible to allow for a transitional period if these rules are to be required
- The pilots indicated that in principle, there is scope to make the EPM more granular than quartiles, provided that the underlying data enable sufficient differentiation between applicants. However many education leads within the medical schools raised concerns that a high level of granularity might be artificial, if assessments are designed to assess minimum competence e.g. pass/fail, rather than to differentiate between the most exceptional and the more satisfactory students. Statistically, increasing the number of bandings can help to differentiate applicants by i) providing a greater spread of scores and ii) reducing the number of applicants with tied scores, without greatly affecting the mean, median or mode. Issues of feasibility and desirability would need to be addressed
- Whilst many schools consult on and publish the composition of quartiles, there is wide variation between medical schools in the way they produce quartiles. The pilot revealed the different compositions of current measures of performance at medical school – which include assessments in all years in some schools to a single assessment in a single year in others. The pilot also raised issues around the comparability of students within a single cohort who had taken a different combination of assessments, for example comparing those on graduate entry with standard entry, as well as the minority of students who transfer between institutions at the end of third year

8.5. EPM Task and Finish Group (EPM Group)
An EPM Task and Finish Group (EPM Group) was convened to recommend specific and workable standards for the EPM framework, in consultation with stakeholders. The EPM Group was Chaired by Professor Mike Greaves, University of Aberdeen, and its membership of twelve included representation from medical school staff, medical students, employers, and the Department of Health. The EPM Group met face to face once in November 2010, with subsequent work carried out by individuals on the EPM Group, and decisions taken electronically. The following chapter outlines the conclusions of the EPM Group, and the process by which these conclusions were reached in consultation with stakeholders.

8.6. Draft framework
A draft EPM Framework was agreed by the EPM Group in February 2011, and proposed to the relevant stakeholder groups (medical school staff, students, employers, ISFP Project Group, ISFP Project Board), and revised to reflect feedback and further recommendations. The final EPM framework was approved by the EPM Task and Finish Group
8.7. Performance in relation to graduating cohort
The EPM Group recommended that the most meaningful comparable measure of performance is of the applicant relative to the graduating cohort at their medical school. There can be no legal requirement for any non-UK medical school to change its curriculum in order to provide a score for application to employment; furthermore, although all UK medical degree courses teach to the GMC’s curriculum framework, Tomorrow’s Doctors, the timing and format of assessments varies considerably between medical schools. Given this, it is neither desirable nor possible to define an EPM framework that stipulates the types of assessment taken in particular years of the degree course.

The EPM Group recommended that the EPM framework should require each school to devise a representative ‘basket of assessments’ covering the curriculum, in consultation with its student body. The standards for the production of EPM scores need to allow for the variety of ways in which assessments are conducted, and must reflect the outcomes to be achieved rather than the means by which they are achieved. In consultation with staff, students and assessment experts, the EPM Group also recommended rules for the consistency of the type of assessments to be included, and the treatment of special cases, for example re-sit assessments and transfer students.

The EPM Group recommended that performance in relation to the graduating cohort be measured in as close to equal sized deciles as possible, allowing for tied scores. The EPM group recognised that the degree of granularity cannot be increased beyond the resolving power of the underlying assessments, particularly given that a number of medical schools record grades rather than raw scores. However, a basket of assessments would provide multi-source performance data, and permit a greater degree of differentiation between applicants. This would be fairer for applicants at the margins of quartiles, whilst not increasing the level of granularity artificially.

8.8. Degrees, presentations, prizes and publications
The EPM Group recommended that a small number of points are available to reward academic excellence, as demonstrated by the possession of a degree, or presentation, prize or publication. The EPM Group accepted the BMA Medical Students Committee’s (BMA MSC) recommendation that these points should be divided into two categories – degrees and then separately presentations, prizes and publications.

There was an extensive consultation with stakeholders around the eligibility of additional points for degrees (whether all previous degrees, only degrees taken within the length of the programme, or only degrees that extended the length of the programme), and the criteria for presentations, prizes and publications. The final recommendations were agreed by all stakeholders, and accepted by the UKFPO Rules Group.

8.9. Extra-curricular activities
The EPM Group recommended that there are no points for extra-curricular activities. The BMA MSC had proposed three possible methods of recording and scoring extra-curricular activities: credits, a standardised portfolio and a free text space to declare additional achievements. However, the EPM Group assessed that the three options were neither cost-effective nor valid, particularly given that there is no evidence that an applicant who has undertaken extra-curricular activities would perform better as an FY1 doctor than any other applicant. Furthermore, the added value of undertaking additional activities and responsibilities (for example through sports, volunteering, committee representation, caring for a relative etc) would already be reflected through performance in the SJT, which assesses the professional attributes of a foundation doctor.
9. Aspects of the application process using SJT and EPM

9.1 Aggregating the SJT and EPM scores
Expert advice was received about how the points for the EPM and the SJT should be aggregated, to provide a total application score. The advice was informed through modelling of score combination techniques, informed by the data of the SJT spring pilots as a basis to estimate the means and standard deviations of key parameters. It was agreed that the aggregation method should provide an appropriate and defensible outcome but also that it should look intuitive, thus gaining a high level of acceptability. This advice concluded that the SJT results should be recalibrated to one decimal point to a 50 point scale and added or averaged with the EPM on a 50 point scale without further data modification. The report of this analysis is provided as Appendix J.

9.2 Algorithm
The report of the ISFP Options Appraisal recommended that there should be further review of the matching algorithm used to match successful and eligible applicants to the Foundation Programme to foundation schools.

Feedback from applicants indicates that the reasons for choosing particular foundation schools are as much a factor of location and personal preference, as they are of the programmes available. Furthermore, patterns of preferences suggest that strategic choices are made by applicants, particularly amongst the London foundation schools which are historically usually oversubscribed.

Applicant preferences, and application scores, are loaded onto a central computer system, and a macro is used to match applicants to foundation schools according to the pre-defined rules. In the event of oversubscription, the highest scoring applicants are selected onto the Foundation Programme, and the remaining applicants placed on a reserve list until such time as a place becomes available.

The matching algorithm used up to FP2011 was ‘First Preference First’, giving priority to applicants for whom a foundation school is their first choice. Applicants are placed into rank order according to their application score. The rank ordered list is then processed by allocating each applicant in turn to her first choice school if a place at that school remains, and if not, deferring further consideration of the applicant to Phase 2. In Phase 2 each applicant in turn is assigned to the foundation school highest on her preference list that still has at least one place remaining.

The alternative matching algorithm considered for selection to the Foundation Programme is known as the ‘Serial Dictator’, which gives priority to applicants with the highest scores first. Applicants are placed into rank order according to their application scores, and each applicant is allocated in turn to the foundation school highest on her preference list that still has at least one place remaining.

Whilst the ‘First Preference First’ algorithm maximises the number of applicants placed in their first choice foundation school, and thus provides greatest applicant satisfaction, it is possible and likely that an applicant A with a low score will be allocated to a given foundation school, even though applicant B may have a higher score, by virtue of their choices. The effect of this is that applicants may select their preferences strategically, and secondly, that applicants may be allocated to a foundation school very low on their list of preferences (once they have missed their first preference) if other schools higher on their list are full. The ‘Serial Dictator’ algorithm achieves a lower allocation to first choice school, but in no case is applicant A, with a lower score, allocated to a foundation school in place of higher scoring applicant B.

Each time a matching algorithm is run, there is random tie breaking for applicants with the same score vying for places in an over-subscribed foundation school. It is worth noting that, because of the rather coarse grained scoring system of the current system, the number of ties that have to be broken is very large, containing several hundred applicants. The
greater granularity provided by SJTs will improve this situation. There is currently a judgement call as to how these tied ranks should be split, as follows:

- The ‘greedy’ rule, which maximises the number of 1st choices, then subject to that, maximises the number of 2nd choices, then the number of 3rd choices, and so on
- The ‘generous’ rule, which minimises the number of last choices, then subject to that, minimises the number of (last-1)th choices, and so on
- The ‘amended-generous’ rule, which maximises the number of 1st choices, then subject to that, minimises the number of last choices, then the number of (last-1)th choices, and so on
- The ‘amended-greedy’ rule which seeks a profile that is the same length as the best possible generous profile, but subject to that, maximises the number of 1st choices, then maximises the number of 2nd choices, and so on

The FP2011 preferences were re-run using the First Preference First and Serial Dictator matching algorithms. The data analysis revealed that with both matching algorithms, the same 5,979 of the 7,073 applicants (85%) would have been allocated to their first choice foundation school. In the First Choice First algorithm, an additional 412 applicants were allocated to their first choice school (90%). Of the 400 applicants allocated to their second choice foundation school through the Serial Dictator algorithm, 134 would have been allocated to their first choice foundation school using the First Choice First algorithm, but 187 applicants would have been worse off, placed between their 3rd and 18th choice foundation school. With both matching algorithms, more applicants in total are placed in the foundation schools of first, second and third choice (92.5% with ‘Serial Dictator’ and 92.3% with ‘First Choice First’). Overall, 412 applicants are better off under the ‘First Choice First’ algorithm, and 458 applicants are better off under the ‘Serial Dictator’ algorithm. However, there is a question of fairness as to which applicants should be allocated to foundation schools lower on their preferences. It should also be considered that the FP2011 data may contain an element of strategic preferences.

The ISFP Project Group reviewed the data, and recommended to the UKFPO Rules Group that the ‘Serial Dictator’ matching algorithm be used to match applicants to foundation schools, on the basis that it was the fairest method by which to do so. At its meeting in May 2011, the UKFPO Rules Group agreed to implement the new matching algorithm, ‘Serial Dictator’, from applications to FP2012, one year before any changes to the selection tools used are implemented.

9.3 Development and testing of SJT item banking software

Since February 2010, The Medical Schools Council Assessment Alliance (MSC-AA) has been developing a new online item bank to store, tag, quality assure and facilitate the delivery of Situation Judgement Test items. Development work has been shared between the MSC-AA Information Systems Team and Epigenesys Ltd., a software development company based in Sheffield. The functional specifications of the system are derived from the work performed by the MSC-AA Information Systems Team on the UKCDR (UK Collaboration for a Digital Repository) specification.

The development methodology adopted was test-driven and feature-driven. The development team generally works on the basis of short iterations of 2 or 3 weeks within which features are designed, implemented and tested in an iterative manner.

A beta version of the system was released in May 2011, and is now extensively used to author and quality assure items. The last large piece of functionality, allowing users to download items directly from the bank, is scheduled to be delivered in October 2011. Since its beta release, the system has been extensively tested in terms of functionality, usability and security. It has now been used in the framework of more than 5 authoring events as well as 20 review events across the UK, and user feedback has been largely positive.
As at 1 August 2011, the system holds a total of 91 SJT items. As additional authoring and quality assurance events are planned in the near future, and as existing content is transferred to the software, this number is expected to rise significantly over the course of the next few months.

**Technology**
The new MSC-AA item bank is a web-based application designed to support the complete quality assurance cycle of the SJT items, from authoring to exam readiness. It also features a multi-level taxonomy tagging system, allowing the tagging of items against an unlimited number of taxonomies.

The application is written in Ruby and makes use of MySQL (relational database management system) and MongoDB (a document-based database) for item storage. JQuery (Javascript library) is also used on the front-end to allow for a rich user experience.

The application is stored on Epigenesys’s servers in Sheffield, and scheduled backups of the database are made on a regular basis. Connection to the server is made through a secure, encrypted, https certificate. Each author and reviewer has unique login credentials to the system, which can be revoked at any time by the system administrators.

**Security**
Penetration testing was carried out on the system by in April 2011 by ProNetExpert Ltd., a security expert company based in Edinburgh. Although security of the system was deemed satisfactory, additional security measures have been implemented to increase the security of the system. Another penetration testing session is scheduled to take place before the end of year 2011, to assess the security level of the application after the implementation of the item delivery feature set.

9.4 **Defence Medical Services**
Each year approximately 60 military cadets, who have been supported by the MoD through medical school, are allocated by the Defense Deanery to supernumerary Foundation Programmes. A memorandum of understanding between the Defence Postgraduate Medical Deanery (DPMD) and the regional deaneries ensures that military foundation doctors are cost neutral, such that the Trust receives the service component as a ‘free good’ and does not charge for the educational component. These students currently do not apply through a competitive process for a Foundation Programme and this process is unlikely to change.

The ISFP Project Group recommends that all applicants to the Foundation Programme should complete the national application process using the SJT and EPM, in addition to any separate processes to recruit to the Defence Medical Services.

9.5 **Academic Foundation Programmes (AFP)**
Around 5% of Foundation Programmes are Academic Foundation Programmes (AFP), which provide foundation doctors with the opportunity to develop research, teaching and leadership/management skills in addition to the competences outlined in the Foundation Programme curriculum. Recruitment to the AFP is currently a regionally coordinated process, usually comprising a standard application form and interview (telephone, video conference or panel), as well as evidence of academic performance at medical school. Applicants are nominated for an AFP by their graduating medical school, and eligibility is checked nationally. Currently, recruitment to an AFP takes place up to six months before selection to the general Foundation Programme. If AFPs are not filled, the posts may be allocated to the general Foundation Programme. Similarly if applicants are unsuccessful, they are still eligible to apply through the national application process. For FP2011, 1,328 applicants applied to 472 AFPs through a two-stage process. There was a single national offers date, followed by a cascade system.
The ISFP Project Group recommends that all applicants to the Foundation Programme should complete the national application process using the SJT and EPM, in addition to any separate processes to recruit to AFPs, before their appointment can be confirmed.

10. Communications
The communications strategy was produced by the Communications Officer for the Medical Schools Council. A number of key areas were addressed in the strategy including the communication objectives and key messages, different stakeholder groups grouped via importance, key activities to help raise awareness of the project, risks, resources and budget. The ISFP Communications Plan is provided as Appendix K.

10.1 Development and sign off
The strategy was developed in April 2010 and circulated to a variety of stakeholders, including the UKFPO, for comments. This was then discussed and approved at a Project Group meeting on 15 April 2010, after which it was presented to Project Board members for final approval and sign off.

10.2 Key achievements
Stakeholder groups were involved in a variety of ways, with information tailored to their needs. All comments from these stakeholders were collated and fed back to the ISFP Project Group throughout the duration of the project. The key achievements are described in sections 10.3 – 10.10.

10.3 Website
An ISFP website was developed to host all relevant information about the project and to help gather feedback from users via an online discussion forum. The website was launched on 5 May 2010 and as at 20 July 2011 there had been 11,675 visits (10,623 from the UK) and 33,351 page loads, with visitors spending an average of 2 minutes 7 seconds on the website. Website visitor numbers varied each month, and the higher levels of visitors coincided with the various communications that were sent out and key project milestones. During the 2010 summer months and over the Christmas period there were very few visits, suggesting that most users were students. Google Analytic data demonstrate that 72% of all visitors came directly to the site, which means they had clicked on the links included in the communications or that they knew the website address. An additional 22% were referred by other websites demonstrating that the weblink which stakeholders put on their website were worthwhile. The most viewed pages were the home page, the SJT main page, and the example SJT question and answers page. The website will continue to be developed during the Parallel Recruitment Exercise (PRE) as the information about the proposed changes develops.

10.4 Social Media
A Facebook Group was set up in May 2010. As at 20 July 2011 there were 1,115 members and 21 comments had been posted on the discussion forum and main page. This has proved an effective way of reaching over 1,000 medical students about the project by sending messages to them at key stages of the project.

A video podcast involving Professor Paul O’Neill and Nick Deakin – who was the BMA MSC representative on the ISFP Project Group at the time – was produced and hosted on YouTube. As at 20 July 2011 there were 1,993 unique views and 5 comments had been posted.

There were 44 comments posted on the ISFP discussion forum between 16 April 2010 and 26 May 2011. The ISFP Project Group responded to some of these as they were either questions or areas that needed to be clarified.

10.5 E-bulletin
An e-bulletin sign up function was added to the ISFP website to allow interested stakeholders to register for project updates. E-bulletins were sent every other month between October 2010 and July 2011 to the 244 registered users.
Updates were also sent via other organisation’s e-bulletins including: UKFPO via its Medical Student Board, Foundation School Directors and Foundation Schools Managers networks, BMA MSC and NHS Employers.

10.6 Communication documents
A number of key documents were sent to medical students, either about the project as a whole or specifically about the SJT pilots. These included: a key facts document to all medical students, a leaflet for all students which was circulated with the Student BMJ and Student BMA News to 20,000 subscribers, a Project Handbook, FAQ documents, letters from Professor Paul O’Neill to all students outlining the project developments in October 2010 and February 2011, and updates for school newsletters. Tailored information was also sent to students at pilot schools including emails about the pilots and workshops with follow up emails two weeks later, and pilot letters to all students outlining what to expect on the day of the pilot and what they needed to bring with them.

10.7 MedSoc Presidents
To ensure that information was reaching as many medical students as possible, the MedSoc Presidents at each medical school were contacted on a number of occasions with information about the project and asked to circulate to all students. Twenty-two MedSoc Presidents replied to confirm they had circulated this information.

10.8 Pilot workshops
Workshops were held at eleven medical schools involved in the pilots between November 2010 and April 2011. The purpose of these was to inform students about the project and to encourage participation in the SJT pilots. Unfortunately, only one of the four schools in the initial pilot was able to host a workshop. This was due to the timing of the pilots and the fact that some students were not back at medical school after the summer period with enough time for this to be delivered. The proportion of medical students who attended the workshop and then attended the pilot was extremely high. As a result, it was agreed that it would be beneficial to hold these workshops at schools involved in the large-scale pilot. Ten schools were able to host a workshop, with the remaining five schools being sent some personalised slides to circulate to all final year students instead. Attendance at the workshops varied due to the timing of these, although feedback was extremely positive. The workshops provided an excellent opportunity to talk through student concerns and gather their feedback on the proposed changes. All schools covered the cost of room hire.

10.9 Presentations
Updates, both oral and written, were provided at a number of key meetings including: UKFPO Medical Student Board, UKFPO Rules Group, UKFPO Careers Annual Conference, BMA Medical Students Committee, Medical Schools Council, MCAN and NEAF (undergraduate and postgraduate careers advisors), GMC Undergraduate Board, Medical Programme Board, UK Scrutiny Group, Medical School Secretaries, Royal Society of Medicine Intercalation Conference and Universities UK staff meetings. NHS Medical Workforce Directors and Medical Directors were also sent email updates.

10.10 Press
Medical press were contacted during each milestone of the project – mainly with information about the pilots. The Student BMA News featured a number of articles, the Student BMJ included two small updates at each pilot stage on their news bite page, the BMJ Careers wrote an article about pre-employment competency checks which mentioned the ISFP project and small number of other medical websites included information which was circulated in press releases.
11. Legal Advice

The arrangements for selecting applicants to the Foundation Programme must comply with the relevant UK and European legislation, notably that relating to employment, equalities, human rights and data protection. Given this, the ISFP Project Group has taken legal advice about the recommendations in this report. The legal opinion is provided as Appendix A.

A briefing for counsel was drafted by the ISFP project team, and agreed with DH. The brief outlined all of the key recommendations relating to the proposed selection methods that had been piloted, summarised as follows:

- Selection will be based on the results of an invigilated SJT and the EPM
- The SJT will be run on a small number of dates in the UK
- Applicants will bear their own travel costs for sitting the SJT
- The EPM will have two components, one reflecting each student's academic ranking in their cohort, expressed in deciles (e.g. the top 10% of the students in the cohort will be in the first decile, the next 10% in the second decile and so on), the other component being awarded for exceptional academic achievements (including additional degrees)
- Each medical school will be required to base its decile framework on a representative basket of academic assessments undertaken by all of the students in a cohort, and to publish the composition of the basket of assessments
- Where applicants cannot provide a decile score, they will be treated as if they were in the lowest decile
- Applicants who re-sit their final year will be assessed on the basis of their original decile (e.g. the decile achieved before the re-sit year), plus any points for additional academic achievements at the time of application
- The professional attributes to be tested by the SJT will be included in the national Person Specification

The legal opinion was sought from a firm of solicitors familiar with the existing arrangements for selection to the Foundation Programme. The opinion was received on 1 July 2011, and reviewed by the ISFP Project Group. In summary, the legal opinion is that the proposed use of the SJT and EPM is consistent with the relevant legislation. Specifically, the solicitors had no concerns about the legality of:

- The inclusion in the national Person Specification of the professional domains identified in the Job Analysis
- The use of an invigilated SJT
- The use of the EPM
- Selecting applicants on the basis of a combined SJT/ EPM score
- The award of additional points for degrees, presentations, prizes and publications, provided that care was taken to ensure that comparable achievements were equally recognised
- Requiring medical schools to publish the basis upon which their EPM scores were determined
- Judging re-sit applicants on the basis of their original decile (e.g. the decile they were awarded in relation to the original cohort before they re-sat a year)
- The possibility of charging applicants to take the SJT
- The possibility of setting a threshold SJT score for the purpose of excluding very low-scoring applicants from selection on the basis of not meeting the national Person Specification

The review has suggested that there remains some small risk of a successful legal challenge in relation to certain equality and diversity considerations, but this appears to be an improvement relative to the risks that are present with the current selection arrangements.

The specific issues raised by the legal opinion are considered in more detail in sections 11.1 – 11.5.

11.1 SJT results of overseas applicants
There was no feasible way to pilot the SJT with a representative sample of overseas applicants during the initial or large-scale pilots. Of the 112 overseas applicants to FP2011 who completed clinical assessments in the UK, only 7 volunteered to participate in the pilot SJT. Members of the ISFP project team travelled to two non-UK medical schools to deliver the SJT pilot, however there were only 43 non-UK participants in total. These data were removed from analysis, so as to not distort interpretation of the results based on a small sample. It will not be possible to compare the performance of UK and non-UK applicants on the SJT until the SJT is used for live selection in FP2013, which will be the first opportunity to obtain data from a representative population of overseas students.

The SJT should at least reduce the risk of a successful challenge compared with the current ‘white space’ questions. In terms of language, the current ‘white space’ questions are a composition test, whereas the SJT is only a comprehension test, which is less onerous.

The ISFP Project Group recommends that the SJT performance of applicants from non-UK medical schools should be analysed after each recruitment round.

11.2 Holding the SJT in the UK
The legal review has suggested that holding the SJT in the UK alone could present a small risk of challenge on the grounds that applicants having to travel to the UK to take the test might be disadvantaged relative to those who live here. The solicitors have recommended that, in order to minimise the risk, the rationale for holding the test in the UK should be properly scrutinised and documented. We agree with this recommendation, and propose that the option of delivering the SJT overseas should be evaluated in more detail. Our current expectation is that delivering paper-based SJTs overseas will not be feasible; however, it may become a viable prospect if the SJT is delivered as an electronic test in the future.

The solicitors have also suggested an investigation of the possibility of setting up a grant scheme to contribute towards the travel expenses of applicants who can be shown suffer financial hardship as a consequence of having to travel to the UK to take the SJT. The ISFP Project Group suggests that this should be taken forward in the run-up to the FP2013 recruitment round. In the meantime it should be noted that the requirement for non-UK applicants to travel to the UK to sit the SJT is no more onerous than the current arrangements requiring eligible non-EU applicants to travel to the UK to sit the clinical assessment as part of eligibility checking for selection to the Foundation Programme.

11.3 Skewing of EPM scores
The current proposals for the EPM require each individual medical school to use a representative ‘basket’ of assessments to rank all of their students in a given cohort. Since the nature and number of assessments varies so much from school to school, the EPM framework cannot specify which individual assessments should be taken into account by any given school. Given this, the solicitors have questioned whether it might be possible for a medical school to base its EPM scores on a deliberately skewed combination of assessments in order to favour its students. Fortunately the nature of the EPM deciles prevents schools from being able to gain an advantage for their cohort. Each applicant is given a score relative to all of the other applicants from the same school, so any change to the ‘basket’ of assessments that favoured one applicant from the school would necessarily disadvantage another from the same school- there is no way to skew the results for the cohort as a whole.

This argument does not apply in the case of a medical school from which only a single student is applying, since in principle it would be possible for the school to skew the selection of assessments to favour that one student while ignoring the effect on all their others. This is an issue that has already been encountered by the UKFPO in relation to the existing academic quartile scores. In practice the scope for such skewing may be reduced by requiring all schools to publish the composition of their EPM basket of assessments and by providing evidence of where the student sits in relation to the whole cohort. It should be noted that this can only currently be required of UK schools.
11.4 Failure to provide a decile score as part of the EPM
Currently, where applicants are unable to provide an academic quartile score they are treated as if they were in the lowest quartile. It is proposed that this arrangement should continue for the decile component of the EPM, so that applicants who do not provide a decile score will be treated as if they were in the lowest decile. The legal advice is that this is a sensible arrangement; however the solicitors have suggested that where applicants can show that they are unable to provide a score through no fault of their own then it would be fairer to provide some other means of judging their academic performance. In practice, however, there appears no workable way of achieving this. Aside from the EPM itself, there is no separate standard test that can be used reliably to equate the academic performance of one applicant compared with a significant sample of others. In principle the grade of an applicant’s final degree, such as 1st, 2.1, 2.2 etc, might be mapped in some way to deciles, but there are two problems with this approach: most medical degrees are not graded as other university degrees are, and in any case most applicants do not sit their finals exams until after they have applied to the Foundation Programme.

11.5 Weighting SJT and EPM scores
It is proposed that the SJT and EPM scores of an applicant should be combined to provide a single overall score for the purpose of selection to the Foundation Programme. This raises the question of the relative weighting to be given to each component to arrive at the overall total. The legal advice is that the weighting given to each component should reflect its relative value as a predictor of the qualities of the applicants; and that whatever weightings are used, they should be published.

12. Management Lessons Learned
One of the purposes of a pilot project is to learn lessons to become better prepared for a live implementation. In this respect, the current project has been very insightful. There has been important learning in the following areas, all of which inform the recommendations in this report to underpin a smooth implementation of the new selection methods:

- Scheduling and managing the delivery of a national SJT in multiple venues
- The practicalities of producing and quality assuring SJT items
- The performance of the SJT as a selection method
- The involvement of volunteers for SJT item development and review
- Paper and electronic delivery options for the SJT
- The variety of assessments across medical schools
- Communications
- Ongoing evaluation
- Recommendation: Parallel Recruitment Exercise (PRE)

These topics are considered in more detail in sections 12.1 – 12.9.

12.1 Scheduling and managing a national SJT
There are two main competing factors to be taken into account in relation to the scheduling of the SJT. On the one hand, many students at UK medical schools spend some part of their final year undertaking an elective period of practical learning overseas; the timing of these ‘electives’ varies from school to school, so that there is no single day in the academic calendar upon which all medical students are in the UK. Requiring large numbers of students to travel back to the UK to take the test during their elective would be costly and disruptive, so this argues in favour of allowing the SJT to be run on a range of dates to suit the timing of electives. One the other hand, there is a significant cost associated with increasing the number of dates on which the SJT may be run, in terms of the provision of venues and also the number of items that would be needed in a single recruitment round. There is a possible risk that students at an earlier sitting of the SJT might leak information about the test to students attending a later sitting. To prevent this, different versions of the SJT test paper need to be produced for each sitting, and the different papers need to be ‘test equated’ to ensure that the overall results take into account any variances in the relative difficulty of the papers.
Aside from the extra cost of producing extra test items, there are other complexities that arise if the SJT is to be scheduled to run on more than one sitting. For example, in order to analyse the outcome of the tests in psychometric terms, it is important that every version of the paper is taken by a sufficiently large and representative sample of applicants, which means that care must be taken to ensure that an adequate number of applicants attends each sitting. While these issues were apparent at the outset of the project, the pilot has provided first-hand experience of their practical implications, which are taken into account in the recommendations.

Other points of learning that have arisen in this area include: the practicalities of distributing and collecting test papers across the UK in a secure and reliable way; the pitfalls associated with the need for each applicant to have a unique identifier that allows their SJT answers to be matched in an unambiguous way to their national Foundation Programme application form; the variation in the nature and capacity of the venues available at medical schools for running assessments; and the time and effort required to produce, quality assure, and mark SJT papers.

12.2 The production of SJT items
The project has adopted a model for the development and quality assurance of SJT items that combines features of two other similar processes, namely the development of academic test items by the MSC-AA and the development of SJT items used by the National Recruitment Office for GP speciality training. The pilot has provided an opportunity to test the process, and gather metrics that will be needed to plan development of SJT items for the longer term. More specifically, the pilot has provided insights including the productivity of ‘item writing workshops’; the proportion of draft SJT items that are suitable for piloting and for live use; the time and effort required to generate and quality assure items; the challenges of maintaining the security of the SJT items.

12.3 The performance of the SJT as a selection method
In all around 1,100 students have taken SJT papers produced during the pilot. This has generated a significant amount of data that have been subjected to psychometric analysis. As a consequence the pilot has yielded important information about the various aspects of the performance of the SJT as a selection method, including:

- The reliability of the SJT
- The time required for applicants to complete the SJT
- The number and type of items to be included within the SJT
- The granularity of the SJT results
- The validity of the SJT

These topics are reported in detail in the report of the SJT pilots (Appendix F).

12.4 The use of volunteers for SJT development
The project has trialled a particular approach to the development and quality assurance of SJT items that is reliant on the participation of volunteers at every stage of the process. More specifically, the volunteers include:

- Clinicians familiar with the role of an FY1 doctor, who draft the SJT items
- Foundation doctors, who review the drafted items to confirm that: the items present scenarios that FY1 doctors might realistically encounter in practice
- ‘Concordance Panel’ members, who sit SJT tests to act as proxies for good applicants

The decision to use volunteers was in part influenced by the MSC-AA model, in which academic questions for use in medical school finals are produced and quality assured by volunteers. This contrasts with the GP model, in which clinicians are paid to draft the SJT items. The experience gained in the pilot has suggested that it is possible to encourage a sufficient amount of voluntary participation overall, but the use of volunteers introduces constraints and leads to unpredictable outcomes.
12.5 Paper and electronic delivery of an SJT

The project has gained experience of three options for delivery of the SJT, namely:

- A paper-based SJT
- An electronically delivered SJT
- A hybrid approach, in which the SJT was completed on paper but applicants used special handsets to record their answers electronically

In each case, the SJTs were delivered under invigilated conditions at venues provided by medical schools. Details about each of the options are presented in section 7.2. The overall learning suggests that the paper-based test is currently the most feasible of the three options trialled. The main reason for this has been, in some medical schools, the lack of suitable facilities for delivering the tests electronically. That said, the creation, distribution, and collection of different versions of paper tests, across a range of venues and dates, is a considerable administrative burden, and presents some challenges in terms of maintaining the physical security of the papers.

Two further options for delivery that could be considered are the electronic delivery of tests either through i) commercially available test venues, or through ii) the use of centrally owned laptops or i-pads. These two methods would avoid the constraints imposed by a lack of infrastructure to deliver assessments to the whole cohort at some medical schools and would overcome the issues of adapting software and access to different hardware and software systems. In any case, if multiple versions of the test are required – rather than for use on common national dates – a larger number of items would be needed and would add to the costs.

12.6 The variety of assessments across medical schools

At the outset of the project there was an expectation that a framework for the EPM could define the type and numbers of assessments that should be taken into account when evaluating student educational performance in relation to cohort. Indeed just such a framework was originally drafted and piloted in the first half of the project. However, this exercise served to highlight the diversity that existed across medical schools in the following areas:

- **Optional v mandatory assessments.** Some schools offered their students a wider choice of optional programmes and assessments
- **Practical v written assessments.** Some schools maintained a sharp distinction between practical and written assessments, while others ran tests that combined both elements
- **Granularity of outcomes.** Some assessments resulted in very detailed scores, while others might result only in a binary pass/fail outcome. Moreover, some schools retained the detailed ‘raw’ scores achieved by their students, while others (as a matter of university policy) retained records in the form of grades (e.g. A, B, C...) and did not retain the raw scores
- **Timing and number of assessments.** Some schools implemented a small number of detailed assessments, while others ran a larger number of smaller assessments. There are also differences in the timing of assessments; some schools tended to hold practical assessments at an earlier stage than other schools

Given the above, it became clear that a prescriptive formula for the construction of an EPM score, of the sort that might dictate a certain mix of practical and written assessments taken from specific years of the curriculum, could not be defined in a way that would provide comparable results across all schools. Instead, it would be necessary for each school to determine which ‘basket of assessments’ would yield the most representative outcome locally.

12.7 Communication

The importance of communications, via a variety of media, must not be underestimated. The key messages must be conveyed to final year medical students taking part in the Parallel Recruitment Exercise (PRE), but also importantly to future cohorts of applicants, both UK and non-UK. It is important that the stakeholder groups aware of the forthcoming changes include employers, deaneries, foundation schools, and clinicians currently involved in marking the white-
space questions – who would in future be involved in SJT item-writing and quality assurance. It will be important to provide the evidence and rationale for the changes, and to ensure familiarity and awareness before the changes are used for live selection to the Foundation Programme.

12.8 Ongoing evaluation
The SJT pilot has demonstrated that the SJT is an appropriate method for selection to the Foundation Programme, as it is able to differentiate sufficiently between applicants, the majority of items performed well and the reliability of the measurement was good.

The commissioning of a new online application process facility provides the opportunity to include a mechanism to facilitate a separate research project to correlate anonymous data collected through selection to the Foundation Programme with past and future performance data, with permissions.

12.9 Recommendation: Parallel Recruitment Exercise (PRE)
Running the pilots has demonstrated the importance of all elements of a live application process being fully piloted before live implementation. Whilst 15 UK medical schools have been involved in the SJT pilots, it is recommended that all 30 UK medical schools with graduating students be involved in a full-scale Parallel Recruitment Exercise of both the SJT and EPM. Legal opinion indicated that participation in the SJT could not be mandated, as it would not ‘count’ towards live application. However, the opportunity to participate in the SJT should be made available to all applicants to FP2012.

A full-scale pilot SJT and EPM provides a unique opportunity:

- To pilot the SJT administrative arrangements in all venues, ahead of live recruitment
- To pilot new SJT items for inclusion in the SJT item bank
- For UK medical schools to consult and publish the ‘basket of the assessments’ to inform the decile component of the EPM score. In the event that an applicant reapplies to FP2013, their original decile score will count – so this will facilitate smooth transition to the new points system in FP2013
- To obtain additional data to evaluate further the use of the SJT and EPM for selection to the Foundation Programme
- To address any administrative or other issues ahead of live implementation

A communications strategy for the PRE has been produced by the Medical Schools Council’s Communications Officer and the ISFP Communications Consultant, and should be followed. This outlines the ways in which students and other stakeholders will be kept informed.

13. Cost Benefit Analysis (CBA)
The decision to develop and pilot the SJT and EPM was based on the outcome of a Cost Benefit Analysis (CBA) which weighed the relative advantages and disadvantages of the SJT and EPM in combination with other alternatives.

13.1 Original CBA
The original CBA was undertaken in the summer of 2009, and weighed the relative advantages and disadvantages of the SJT and EPM compared with other competing options that had been identified by an expert panel. The other options considered during the Options Appraisal are summarised in Figure 17.
Improving Selection to the Foundation Programme

Final Report

<table>
<thead>
<tr>
<th>Option Name</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do Nothing</td>
<td>Continuing with the current approach to selection, retaining the ‘white space’ questions and academic quartiles</td>
</tr>
<tr>
<td>Structured Interviews/ EPM</td>
<td>Each applicant would meet with a panel of interviewers and be asked a defined set of questions with an agreed, structured scoring system. The scored interview would replace the use of ‘white space’ questions. There would continue to be a need for medical schools to provide something similar to the academic quartile score to take into account the applicant’s academic performance</td>
</tr>
<tr>
<td>Multiple Mini Interviews (MMI)/ EPM</td>
<td>MMIs are an extension of structured interviews, in which the applicant rotates around a series of short interviews with each interview being designed to explore a particular element of the personal specification for foundation training. Each interview is structured with a calibrated scoring system. Again the scored MMIs would replaced the use of ‘white space’ questions, and medical schools would continue to provide something similar to the academic quartile score to take into account the applicant’s academic performance</td>
</tr>
<tr>
<td>National Assessment</td>
<td>Every applicant would undertake a standard assessment (which is separate from their Finals) of their clinical knowledge and skills. The assessment would replace the use of the ‘white space’ questions and would obviate the need for a separate academic score from medical schools</td>
</tr>
</tbody>
</table>

Figure 17: Alternative selection tools considered during Options Appraisal

The CBA followed the recommendations of the Treasury’s ‘Green Book’ which provides a framework for the formal evaluation of options. Estimates were made of the costs of each option over five years based on rates current in 2009, with a 3.5% annual discount factor. The relative costs of the options were estimated as in Figure 18.

<table>
<thead>
<tr>
<th>Option</th>
<th>Five year cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do Nothing</td>
<td>£21m</td>
</tr>
<tr>
<td>Structured Interview/ EPM</td>
<td>£23m</td>
</tr>
<tr>
<td>MM1/ EPM</td>
<td>£26m</td>
</tr>
<tr>
<td>National Assessment</td>
<td>£32m</td>
</tr>
<tr>
<td>SJT/ EPM</td>
<td>£19m</td>
</tr>
</tbody>
</table>

Figure 18: Relative costs of selection tools in Options Appraisal

To evaluate the non-financial pros and cons, the options were scored against a suitable set of evaluation criteria, shown in Figure 19.

<table>
<thead>
<tr>
<th>Criterion Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>The technical reliability of the selection technique associated with the option. Broadly this means the likelihood that applying the technique in the right way will give consistent results.</td>
</tr>
<tr>
<td>Validity</td>
<td>The technical validity of the selection technique associated with the option. Broadly this means the extent to which the technique is actually measuring the characteristics of a ‘good’ doctor.</td>
</tr>
<tr>
<td>Granularity</td>
<td>The extent to which the selection method provides a fine-grained set of scores for the applicants.</td>
</tr>
<tr>
<td>Consistency</td>
<td>The extent to which the selection techniques can be expected to be consistently applied across the UK.</td>
</tr>
<tr>
<td>Longevity</td>
<td>The extent to which the performance of the option can be maintained over successive recruitment rounds.</td>
</tr>
</tbody>
</table>
The options were scored against each of the criteria on a scale of 0 to 10, where 0 represented the worst option in terms of a given criterion, and 10 the best. The criteria were then each given a weighting to take into account their relative importance, and the significance of the spread represented by the scores from 0 to 10. The results of this are summarised in Figure 20.

A sensitivity analysis showed that the combination of SJT/ EPM had the lowest cost and highest score regardless of any realistic changes to the underlying assumptions, and so should be the preferred option on both counts.
13.2 Updated CBA
The CBA has been revisited in the light of the experience gathered during the project. Most of the assumptions made in the original cost benefit analysis have been shown to remain valid. The two main factors needing revision are costs and feasibility, considered in section 13.3 and 13.4.

13.3 Costs
A key assumption underlying the original cost estimates was that the all-in cost of developing, maintaining and delivering the SJT would be the same as the costs incurred by the GMC in developing, maintaining and delivering Part A of the PLAB (Professional and Linguistics Assessment Board) test, which is used to assess doctors who have qualified in medicine outside of the EEA. The PLAB test appeared to be a reasonable comparator on the grounds that, like the SJT, it was an invigilated, machine-markable, multiple-choice test, for which the items were developed by clinicians. The PLAB Part A fees at the time were £145 per applicant, which equated to £1.13m per year for an SJT to be taken by 8,000 applicants.

The deciles component of the EPM will be maintained by individual medical schools, which is at no cost to the national delivery of the recruitment process; although as with the current system of quartiles, does require significant staff resource. There will need to be a method of verification for points for additional academic achievements.

Current estimates suggest that the costs for the use of the SJT and EPM in combination for selection to the Foundation Programme will be lower than first expected, being nearer to £0.9m per year, rather than £1.13m per year. This sum excludes the running costs of the UKFPO and FPAS. The costs cover:

- The secure storage of SJT items
- Maintaining a bank of quality-assured SJT items to support the year-on-year requirements for selection to the Foundation Programme. This is to include:
  - Analysing the performance of items
  - Removing or improving items with inadequate performance characteristics
  - Updating items that may have become dated as a result of changes to terminology, standards etc
  - Forecasting the demand for new items
  - Creating and quality assuring new items to meet the projected demand, including the piloting of new items
- Defining the standards for the arrangement, conduct and invigilation of SJTs by medical schools
- Providing suitable guidance about the standards, and ensuring that the standards are understood, and adhered to, by medical schools
- Creating, printing, distributing, delivering, collecting and marking SJT papers (assuming the test is delivered on paper), and providing the results of the test to UKFPO in an agreed format
- Arranging, hosting, invigilating and administering SJTs at venues throughout the UK
- The maintenance and evaluation of the EPM framework and standards
- Providing suitable guidance about the EPM framework, and ensuring that the standards are understood, and adhered to, by UK medical schools
- Promoting awareness and understanding of the SJT and EPM
- Providing decile points for UK applicants, as part of the EPM
- Verification of decile points for non-UK applicants, as part of the EPM
- Verification of points for degrees, presentations, prizes and presentations within the EPM
13.4 Feasibility
Given that the SJT/EPM option has been piloted, and the other options have not, this is now a more feasible proposition than it was when the original CBA was performed. Therefore the score for the feasibility of SJT/EPM needs to be increased relative to the other options, which strengthens the case for SJT/EPM in terms of the financial and the non-financial factors.

13.5 The case for change revisited
Figure 21 summarises the case for change; it lists the main concerns about the current selection process, and indicates the corresponding improvements that can be expected to be brought about by the proposed changes.

<table>
<thead>
<tr>
<th>Concern about current system</th>
<th>SJT and EPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>The use of ‘white space’ questions is not sustainable as a selection tool and will become steadily less discriminatory between eligible applicants given that i) there is a limited range of new questions that can be generated</td>
<td>SJTs will draw upon a bank of items to be available for each application round. Given that the situations experienced in the Foundation Programme are varied and complex, new items can be built incrementally and continuously against the detailed Job Analysis</td>
</tr>
<tr>
<td>The marking of ‘white space’ questions is labour intensive, the cost in clinician time approaching £2 m per year</td>
<td>SJTs are machine markable. The clinician time involved in the development and quality assurance of SJT items is expected to be between £50,000 and £100,000 annually</td>
</tr>
<tr>
<td>The use of ‘white space’ questions in non-invigilated conditions – and the availability of model answers on the internet – raises concerns about the risk of plagiarism and coaching</td>
<td>SJTs will be undertaken in invigilated conditions in the UK on no more than three national dates It is not possible to revise for the SJT as the scenarios are complex and answers relate to judgement rather than knowledge</td>
</tr>
<tr>
<td>While the ‘white space’ questions offer a practical way to rank large numbers of comparable applicants, their technical reliability and validity is open to question</td>
<td>There is evidence for the validity of SJTs dating back more than 30 years. The SJT pilots demonstrate the technical reliability, internal reliability, and validity for use for selection to the Foundation Programme</td>
</tr>
<tr>
<td>The academic quartile system makes it difficult to compare fairly between applicants from different medical schools, as it is not standardised or subject to quality assurance across medical schools</td>
<td>The EPM includes a standardised framework for deciles, within which medical schools have the flexibility to define in consultation with their student body, the contents of the ‘basket of assessments’. Schools will be required to publish their locally agreed deciles framework, which will facilitate transparency and quality assurance from the wider community. The move to deciles will also be fairer to applicants at the margins</td>
</tr>
<tr>
<td>If there is an increase in the number of eligible non-UK applicants such that the Foundation Programme is over-subscribed, the selection process may be more likely to be subject to legal challenge from unsuccessful applicants to the Foundation Programme</td>
<td>The use of a valid SJT in invigilated conditions, and EPM in combination are a more reliable, robust, fair and valid method of selection to the Foundation Programme, and thus likely to withstand legal challenge</td>
</tr>
</tbody>
</table>

Figure 21: The case for change revisited
14. The Way Forward

14.1 Final recommendations

- Selection to the Foundation Programme should reflect the skills, knowledge and professional behaviours of the applicant, reflecting the integrated nature of the Foundation Programme as both education and employment
- Selection to the Foundation Programme from FP2013 onwards should be based upon:
  - An invigilated Situational Judgement Test (SJT) to assess aptitude for the Foundation Programme (to replace ‘white space’ questions); and
  - An Educational Performance Measure (EPM) to reflect educational performance at medical school up to the point of application to the Foundation Programme (to replace quartiles)
- Other aspects of the process of application to the Foundation Programme should remain unchanged, namely a national application process including applicant declaration of foundation school preferences, the use of a matching algorithm and eligibility checking
- There should be a full-scale shadow Parallel Recruitment Exercise (PRE) SJT and EPM run concurrently with the current application process for FP2012
- Using data from tracking, there should be ongoing evaluation of the validity of the SJT and EPM as methods for selection to the Foundation Programme

Assuming the final recommendations of the ISFP project are accepted, more detailed arrangements to operationalise the implementation for FP2013 are set out below. These will also be subject to ongoing evaluation and can be adapted as appropriate for future recruitment rounds.

14.2 Operational Guidance for FP2013

- Scores for performance on the SJT and EPM should be equally weighted
- A matching algorithm, as agreed by the UKFPO Rules Group, should be used to match applicants to foundation schools according to their preferences and their application scores
- Matching of applicants to individual programmes should remain a local process, managed by the foundation school to which the applicant is allocated
- All applicants to the Foundation Programme, including applicants to academic and to the defence medical services, should complete an SJT and EPM. Separate methods of recruitment to these programmes should continue to run in parallel
- In the event of oversubscription, the highest scoring applicants should be selected to the Foundation Programme. The remaining eligible applicants should be allocated to Foundation Programme posts as vacancies arise, according to their application score
- On application, applicants should provide consent that their subsequent performance will be tracked anonymously
- The communications strategies for both the PRE and the live application process should be implemented

14.2.1 SJT Operational Guidance

- The SJT should consist of 60 live questions. Around two thirds of the questions should be in the format ‘rank five options’ and one third of the questions in the format ‘select three from eight options’
- The content of the SJT should be mapped against the professional domains identified in the Job Analysis
- The SJT should be taken under invigilated examination conditions in the UK, on no more than two nationally agreed dates between November and January. An additional date should be offered to those applicants with extenuating circumstances. The SJT should also be offered at the same venue as the clinical assessments undertaken as part of eligibility checking
• A single SJT question paper should be used on any one date and different versions of the test used for test equating purposes
• Applicants may only take the SJT once during a single recruitment round
• There should be national guidance for the standards of delivery of the SJT by test centres, and a Service Level Agreement in place to require that these standards are adhered to
• There should be agreed national rules for e.g. eligibility of extenuating circumstances, provision for special requirements
• UK medical schools should be responsible for delivering the SJT to their graduating cohort, in line with national guidelines
• A separate venue (central location) should be booked for the delivery of the SJT to non-UK applicants, in line with national standards
• The SJT should be paper-based and machine-markable in the first instance, with further consideration of computerised delivery of the SJT in future years
• There should be centralised printing, scanning of answer sheets and marking
• The Quality Assurance processes recommended by the review of the SJT pilots should be followed.
• The scoring convention used for the pilots should be used for FP2013
• Only SJT items with confirmed acceptable psychometric properties through post-test analysis will count towards the applicant’s SJT score
• There should be a programme for developing new SJT items. Alternative models of item-writing should be trialled and reviewed for effectiveness
• New items should be piloted in addition to the SJT live test content within the SJT papers, in order to develop the SJT item bank for use in live selection
• Secure SJT item-banking arrangements should be in place
• Further work should be undertaken to explore whether candidates who perform extremely poorly in the SJT do or do not meet the national Person Specification.

14.2.2 EPM Operational Guidance
• The EPM score should be derived from educational achievements in three parts:
  • Performance at medical school in relation to the final year cohort, banded by deciles
  • Additional degrees (Bachelors, Masters and Doctorates)
  • Academic achievements including presentations, prizes and publications
• All UK medical schools should publish the composition of the framework by which the decile points are determined, informed by local consultation, according to the principles of the EPM framework
• UK medical schools should provide a decile score to the UKFPO in an agreed format
• Non-UK applicants should submit a Dean’s Statement confirming i) the size of the graduating cohort, ii) their class rank within the graduating cohort and iii) the decile points awarded
• Where an applicant is unable to provide a valid decile point score, they will receive the same number of points as the lowest decile
• Applicants should provide evidence of degrees, presentations, prizes and publications, to be eligible for these points. A mechanism for verification of the evidence should be put in place
15. Success Criteria
The success criteria for the ISFP Project are as follows:

- **Completeness.** All of the project’s products have been completed and accepted
- **Legality.** Legal advice has confirmed that the proposed approach is robust
- **Finances.** The projected costs of implementing the new approach are less than the costs of the current process
  The actual costs of the project are within the approved budget
- **Reliability and Validity.** The tests have confirmed that SJT and EPM provide a reliable and valid selection method
- **Practicality.** The project has confirmed that: it is possible to schedule and run the SJT for the expected numbers of applicants, taking into account the need to avoid electives; medical schools are able to provide EPM scores to the required standard, and there is a satisfactory means of dealing with applicants that cannot provide EPM scores
- **Buy-in.** The key stakeholders have confirmed their support for the approach, including:
  - The 4 UK Health Departments
  - All foundation schools
  - All medical schools
  - BMA Medical Students Committee
  - UKFPO
  - COPMED
  - General Medical Council
  - NHS Employers
  - Medical Education England
  - Medical Royal Colleges
- **Robustness.** Piloting has shown that it is possible for medical schools to implement the SJT securely, taking into account the need to keep questions secret, to ensure the identity of the applicants taking the tests, and to mark the scripts correctly
- **Timeliness.** The key products have been delivered in good time for the start of the application process for FP2013
- **Sustainability.** It is possible to develop a sufficiently large number of SJT items over many years