Foundation Programme Steering Group

Cost Benefit Analysis of Options for Selection into Foundation Programme

August 2009

Contents

1. Introduction 53
2. Context 54
3. The Problem 55
4. The Approach 58
Summary of the Method 58
Valued Costs and Benefits 59
Unvalued Costs and Benefits 59
Sensitivity analysis 59
Risks 59
Practical Approach 60
5. Options Considered 61
Common Elements 62
Do Nothing 63
Multiple Mini Interviews 63
Structured Interviews 64
National Assessment for Ranking 64
Educational Performance 65
Situational Judgement Test 66
Matching Rules 66
6. Assessment Criteria 67
7. Common Costing Factors. 69
Scope and Timescales 70
Applicants and Posts 70
Schools 70
Central Administration and Allocation System 70
Procurement 71
Manpower 71
Piloting and Development 72

Common Content of Assessment 72
Legal Challenge 72
NPV Calculations 72
8. Monetary Analysis 74
Do Nothing 74
Multiple Mini Interviews 76
Structured Interviews 78
National Assessment for Ranking 79
Situational Judgement Test 80
Educational Performance 81
Summary 82
9. Non-Monetary Analysis. 82
Reliability 83
Validity 83
Granularity 84
Consistency 85
Longevity 86
Educational Impact 86
Fairness 87
Compliance 88
Transparency 88
Applicant Burden 89
Medical Time 90
Feasibility 90
Public Opinion 91
Weightings 92
Weighted Performance 93
10. Comparisons and Sensitivity Analysis 94
Initial comparison 94
Cost Insensitivities 95

Dominant Cost Factors 95
Interviewer Costs 96
Unit Costs for Assessments 96
CCA costs 96
Scoring Uncertainties 96
Weightings 97
Optimism Bias and Risk 97
Overall NPV Uncertainties 98
Conclusion 99
11. Document History 101
12. Annex A- Alternative Weightings 102
DH Weightings 103
FRG Weightings 103
Stakeholder Meeting Weightings 1 104
Stakeholder Meeting Weightings 2 105

1. Introduction

1.1. This document is a report on the cost benefit analysis (CBA) work that has been completed for the Foundation Programme Steering Group (FPSG) as part of their assessment of options for the recruitment of trainee doctors to the first year of the Foundation Programme (F1) across the United Kingdom.

1.2. The document has the following structure:

Context. Describes the role and work of the FPSG in order to provide a context for the remainder of the report.

The Problem. Summarises the issues with the current approach to the recruitment of F1 doctors, and the main factors that might identify a suitable replacement.

The Approach. Describes the practical and theoretical approach to the CBA work.

The Options. Summarises each of the options that have been considered.

The Criteria. Defines the criteria that have been used to select the options.

Common Cost Factors. Describes the standard financial assumptions that underlie all of the costings of options.

Monetary Analysis. Presents estimates of the overall value of each option in monetary terms.

Non-Monetary Analysis. Presents the relative performance of each of the options in relation to the non-monetary assessment criteria.

Inter-Comparison and Sensitivity Analysis. Considers the overall attractiveness of the options, taking the monetary and non-monetary factors into account. Considers the extent to which likely changes to the assumptions made in the analysis could materially affect the relative attractiveness of the options considered, and the possibility of combining options.

Annex A. Provides details of alternative sets of weightings that have been suggested by stakeholders in relation to the assessment criteria.

2. Context

2.1. This section provides a brief description of the role of the Foundation Programme Steering Group (FPSG), insofar as is needed to provide a context for the remainder of the document. This section may be skipped by readers familiar with the FPSG.

2.2. Since 2006 a standard UK-wide process has been followed to recruit new medical graduates into their first year of post-graduate training (known as the F1 year). As a result of concerns about the process (see Section 3) the FPSG has been established to lead a review, commissioned by the Department of Health in England (DH), to determine whether and how a more robust process could be implemented.

2.3. The FPSG has commissioned three literature reviews to identify the available evidence relating to the reliability and validity of selection methods for medical recruitment. The FPSG has also commissioned an Expert Panel to review the evidence, and to make qualitative assessments of possible options for F1 recruitment.

2.4. The work covered by the current report is a formal cost benefit analysis of the options identified as feasible by the Expert Panel. The purpose of the work is to provide the FPSG with quantified information about the relative advantages and disadvantages of each of the options, in order to support the FPSG's decision making.

2.5. The terms of reference for the current work are as follows:

- To work proactively with the Expert Panel to ensure that all evidence presented to the Steering Group is consistent, and as well informed as possible. This will be facilitated by the Secretariat [of the Medical Schools Council].
- Translate the evidence from the literature reviews, the extensive consultation with stakeholders, any other appropriate sources, and the qualitative assessment of the Expert Panel into a robust cost benefit analysis of all viable options specified by the Expert Panel, on which the Steering Group can make a recommendation of the preferred option.
- Follow best practice guidance in option appraisal, one source being HM Treasury's Green Book.
- Consider the full range of costs, benefits and risks for each option (identified and detailed by the Expert Panel) in the context of: the current system, the project objectives as outlined in the Dec 15th proposal to DH, and from the perspective of all stakeholders.
- Identify any conflict of interest e.g. service provider, employer.
- Provide an account of how the information assessed was evaluated and conclusions drawn.
- Assist the Steering Group in engaging with stakeholders to explain the evidence and rationale behind the recommendations and to reflect upon the reactions and consider any necessary adjustments.

3. The Problem

3.1. Every year several thousand medical students graduate from UK medical schools to take up medical training posts. In the UK the first two years of post-graduate medical training is delivered through a scheme known as the Foundation Programme; junior doctors work in posts at hospitals and receive educational development from a Foundation School, which is usually part of the local medical deanery.

3.2. Since 2006 there has been a UK-wide process through which students have applied for post-graduate training positions within the Foundation Programme. The process has two broad stages. In the first stage the applicants are allocated to individual Foundation Schools; in the second stage they are allocated to individual jobs associated with their allocated school. The current work is focused mainly on the options for the first of these stages (although, as seen in later sections, the attractiveness of an option for the first stage might depend on the extent to which it facilitates the second).

3.3. For the purposes of allocation to a Foundation School, applicants complete an online form in which they answer 'white space' questions designed to test the quality of their application. The answers to the white space questions are scored according to a nationally agreed set of scoring rules. The applicants also receive a less granular score from their medical school, which ranks them in a given quartile of all the students in their year. A weighted combination of the application form score, and the medical school quartile score, is used to determine an overall score for each applicant. A set of rules (a 'matching algorithm') is then applied to the overall score to determine which applicant goes to which Foundation School, this being driven by the applicants' stated preferences and the competition for available places[1].

3.4. The process follows a standard national timetable. Applicants may submit their application at any time during an application period, the length of which has varied from year to year but has typically been a few weeks starting in the autumn of the applicants' last academic year. The results of the allocation to Foundation School are announced early in the following calendar year.

3.5. The main issues that have led to this review of the current process are as follows:

- 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 work of the applicant; and 'model' answers have been offered for sale. There have been concerns that the impact of such model answers is likely to increase with time, as it will become increasingly difficult to invent entirely fresh questions for each successive recruitment round.
- There have been questions about the extent to which the academic quartile scores can be equated across medical schools.
- While the white space questions appear to offer a practical way to rank large numbers of comparable applicants, it is not clear that they actually measure the qualities of a 'good' applicant (leaving aside the issue of what 'good' means in this context).
- The marking of the white space questions is labour intensive.

3.6. The literature reviews commissioned by the FPSG have added to these concerns, as there is a lack of published evidence to support the use of white space questions as a legitimate basis for selection.

[1] There are approved exceptions to the process outlined above for special cases, such as applicants who are members of the armed forces, and those with special needs.

3.7. To date the impact of these concerns has been softened considerably by the fact that there have been more posts than applicants, and around 90% of applicants are allocated to their first preference Foundation School. From this perspective the process has generally been seen as positive by those involved. However, the perspective is somewhat misleading. The percentage of applicants who get their first choice is a product of the algorithm that is used, and the relative demand for places at schools- it is not dependent on the selection or scoring method.

3.8. Furthermore, there is no guarantee that the number of posts will continue to exceed the number of applicants. The effect of European employment law is that the F1 jobs must be offered on an equal footing to applicants from the European Union; and in any event applicants from elsewhere can apply subject to the prevailing 'right to work' arrangements. In the event that the number of applicants did exceed the number of posts, then it would be more important that the process for selection did indeed pick the 'best' applicants from the pool available.

3.9. In considering whether a more suitable approach might be implemented, several competing factors must be taken into account, such as:

Educational Impact. Some selection methods will support educational objectives more than others. For example, concerns have been raised that some options might put unhelpful pressure on medical schools to change their curricula, or distract students from the completion of their studies.

Disruption to the NHS. Some people suggest it is important for medical staff, who will need to work with the foundation doctors, to have more say in selection. However, a selection method that called for significant participation by medical staff across the NHS would have a high cost of delivery, distracting skilled people from service delivery.

Reliability and validity. Some techniques for evaluating applicants- such as invigilated examinations- can be relied upon more than others to deliver consistent results; however, the techniques that are most reliable are not necessarily valid tests for all of the qualities that make the 'best' doctors.

Timing. The selection of doctors has to be made on the basis of information that is available at the time. To date the applications have been made several months before the students complete their finals, so the allocations to Foundation School takes place before it is known whether every applicant will qualify. More importantly, the medical schools differ in the way in which they deliver their curricula. In one medical school, for example, the students might cover in their fourth year of studies topics that are not covered in other schools until the final year. This means that care must be taken to base the selection on factors that are common to all students at the time, otherwise the process would be both unfair and invalid. A further constraint with the timing of the process is that medical students spend periods of time studying away from their medical school in temporary placements known as 'electives', often in other countries. It is difficult to schedule a process in a way that does not clash with the timing of the electives of a significant proportion of applicants.

Uncertainty. There is no perfect way of determining whether one prospective doctor will be 'better' than another- at best a selection method might make predictions with some statistical validity and a good deal of uncertainty. It is important to recognise that these factors of uncertainty are large for all selection methods, so a method that will always pick the 'best' doctors is an unattainable goal.

Consensus. Finally, even if there is some basis upon which one selection method can be judged to be preferable to the others, it is likely that the judgement will have to be made by balancing the relative importance of competing interests across different parts of the stakeholder community, which could lead to a loss of good will. Even where no hard compromises have to be made, there is still a need to get all stakeholders to understand and respect the reasons why a particular decision has been taken. Where the factors involved in the decision are complex, which is certainly the case in deciding upon an approach to F1 recruitment, a significant communications effort will be required to justify the decision to the satisfaction of the stakeholders.

3.10. The task of the current work is to consider and describe the beneficial and adverse impacts of the options in a systematic way, quantifying them in monetary terms where possible, to present the FPSG with a clear inter-comparison of the options in order to inform their decision making. The approach to the work is defined in the next section.

4. The Approach

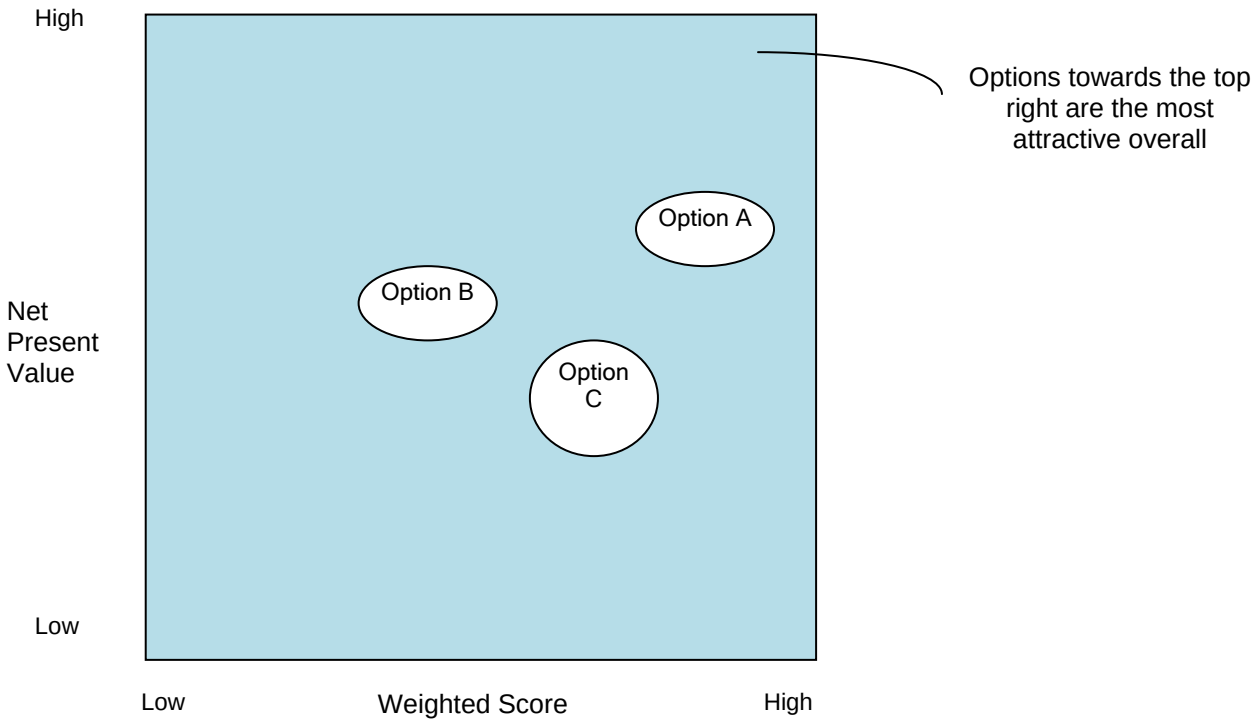
4.1. This section describes the method that has been used for the current work, and provides a brief summary of the practical aspects of its implementation.

Summary of the Method

4.2. A wide range of methods exists for the formal evaluation of options. For the present work several constraints and considerations narrowed the choice. In particular, the option appraisal needs to satisfy the Government requirements for a 'policy impact analysis', which in turn means that it should be compliant with the official guidance published by H.M. Treasury, known as the 'Green Book',

4.3. The Green Book distinguishes between 'valued' and 'unvalued' costs and benefits: i.e. those that can be given a reliable monetary value, and those that cannot. For the former, the guidance requires a net present value (NPV)[2] to be calculated for each of the options, using CBA techniques; for the latter the guidance suggests options should be scored against assessment criteria.

4.4. Given the above, the relative attractiveness of the options can be seen in a straightforward way, as shown in the simplified example below:



[2] A net present value is a particular way of expressing the overall cost (or saving) associated with an option.

4.5. In the example shown above, Option A dominates the others both on value and on score. Where the outcome is not so clear cut, other methods of ranking will be considered, such as benefit/cost ratios, which can be determined both for valued and unvalued benefits.

4.6. The method is described in rather more detail below.

Valued Costs and Benefits

4.7. Where reliable figures are available, the costs and benefits of an option will be quantified in monetary terms and summarised as a net present value (NPV) for the option. Each element of cost and benefit will be expressed as an expected value and an associated range of uncertainty. The costs and benefits will be modelled over 5 years using appropriate discount factors.

Unvalued Costs and Benefits

4.8. A set of assessment criteria C_i is defined to represent the desired characteristics of an F1 selection process, against which each option O_j can be given a score s_{ij}

4.9. For any criterion the scores should range from 0 to 10, where 10 will be allocated to the option that best meets the criterion, and 0 to the worst (i.e. 'local' scoring, rather than 'global').

4.10. Each criterion is allocated a weighting w_i.

4.11. A performance matrix is constructed, containing the scores, s_{ij} of each option against each of the assessment criteria.

4.12. Each option achieves an overall weighted score S_j where:

$$Sj = \Sigma_i W_i S_{ij}$$

Sensitivity analysis

4.13. The options are ordered by NPV and by overall score. The NPVs and overall scores are subject to sensitivity analysis to identify whether there are any likely variations in scores, weightings, or in the estimated values of costs and benefits that could reverse the relative ranking of pairs of options.

Risks

4.14. Risks may be accounted for within the method in any of three ways as follows:

As a single assessment criterion. This is a broad brush approach where risk is quantified for each option as a score against a single assessment criterion for risk.

Within the other assessment criteria. Where a risk can be associated with an assessment criterion, then the variation in risk across the options can be taken into account when scoring the options against that criterion.

Through the sensitivity analysis. Where a risk equates to an uncertainty in a given score, cost or benefit, then the impact of the risk can be considered through adjusting the corresponding values during the sensitivity analysis.

Adjustments to the NPV. Where there is a basket of risks that can be meaningfully given an expectation value in monetary terms, then they can be treated as additional costs in the calculation of the NPV for an option.

Practical Approach

4.15. The current work started on 12th May 2009. The first activities were: to liaise with the Expert Panel to understand the likely scope of the options to be considered; to define in detail the analytical method to be used; and to review the material available from the literature reviews and from the earlier stakeholder meetings and questionnaires organised by the FPSG.

4.16. A formal description of the proposed method was drafted and circulated for initial quality assurance on 19th May 2009, then tabled at the FPSG meeting on 21st May.

4.17. At the same meeting, the FPSG requested that DH and the CBA team should draft a set of assessment criteria for it to approve (see Section 6). The draft was circulated on 1st June 2009 having been derived from an analysis of the feedback from the earlier stakeholder meetings.

4.18. The Expert Panel delivered its conclusions at the FPSG meeting on 29th May, following which the CBA team was asked to liaise with the chair of the Expert Panel to develop the set of selection options to be modelled in the analysis. The options were set out in a 'qualitative cost model' distributed for FPSG approval on 1st June. This model identified the cost elements associated with each of the options, and proposed how those costs would be estimated for the purpose of the CBA work.

4.19. A draft version of this report was distributed to the FPSG on 8 June 2009, with scores and weightings that had been drafted with the help of two FPSG members. Three further versions were produced following feedback from FPSG members, the current version taking into account the outcome of discussions at the FPSG meeting on 26th June 2009.

5. Options Considered

5.1. This section introduces and describes the options that have been considered within the analysis. The descriptions are designed to be sufficiently detailed to allow the analysis to be understood. Readers who wish to understand the background to the options in more detail are referred to the final report of the Expert Panel, and the minutes of the FPSG meeting of 29th May 2009.

5.2. Each of the options presented here is based on a particular selection technique. In principle it would be possible to implement combinations of these techniques in order to increase the validity of the selection process. Given this, the options presented here should be considered as building blocks- which the FPSG may wish to re-combine to present as options for their final report- rather than an exhaustive list of the possible combinations.

5.3. The 'building block' options are summarised in the table below, then individually described in more detail after a description of the elements they share in common.

Name	Summary
Do Nothing	This is the retention of the current approach to selection.
Structured Interviews	Each applicant meets with a panel of interviewers and is asked a defined set of questions with an agreed, structured scoring system. The interview score is combined with the educational performance score for the purpose of selection and allocation.
Multiple Mini Interviews (MMI)	Multiple Mini Interviews are an extension of structured interviews, in which the applicant rotates around a series of short interviews with each 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 interview score is combined with the educational performance score for the purpose of selection and allocation.
National Assessment for Ranking	Every applicant undertakes a standard assessment test (which is separate from their finals) which considers their knowledge and clinical skills. The score achieved by the applicant is used for selection.
Situational Judgement Test (SJT)	Every applicant undergoes an invigilated test, which is designed to assess their professional judgement and likely behaviours in scenarios based on complex situations encountered in Foundation training. The scores achieved from the tests are combined with the educational performance score for the purpose of selection and allocation.
Educational Performance	In this option, fine grain education performance scores are provided by the medical schools according to an agreed common framework based primarily on assessments of applied knowledge and clinical skills.

5.4. The report also considers a secondary option to change the matching algorithm that is used for allocating successful applicants to foundation schools.

Common Elements

5.5. Before describing the options in detail, it is useful to describe the characteristics that they share, so that these common elements can be excluded from the individual descriptions.

The most important of these common elements is the use of an educational performance score, which occurs in all of the options bar National Assessment and Do Nothing. These scores will be provided by the medical schools for each of their students, and will be determined in a standard and open way according to nationally agreed guidelines. The scores will be primarily based on assessments of applied knowledge and clinical skills, but may include performance in other elements if agreed by medical schools for inclusion within the national framework.

5.6. The educational performance scores will address one of the main concerns about the current selection method, namely that it assumes that academic quartiles can be equated across medical schools. This is an understandable concern, more so in connection with medical schools from outside the UK[3].

5.7. Some medical schools in the UK have been cooperating to develop shared banks of assessment items, based on the educational outcomes defined in the GMC publication 'Tomorrow's Doctors'. The proposal is that medical schools will draw items from a common bank to form at least a minimum percentage of their assessments in the normal course of their educational programmes, thus providing additional assurance about common levels of quality in educational delivery. The combination of items used in any one assessment will differ between schools, but the bank will contain both application of knowledge items and also benchmarked Objective Structured Clinical Examination (OSCE) stations based on important concerns for NHS employers such as patient safety and communication. In evaluating the options suggested by the Expert Panel, it is assumed that this 'Common Content of Assessment' initiative will be taken forward by all UK medical schools for reasons of educational improvements, regardless of whether or not the CCA has a role to play in selection.

5.8. The remaining common elements are characteristics of the current system that are expected to be carried into all the options, as follows:

5.9. **Merit and Scoring.** All of the options are based on the principle that the 'best' applicants should be selected. It is assumed that all eligible applicants will continue to be scored to a common standard across the UK, and this scoring will be used both for allocation and for selection (in other words, if the number of eligible applicants exceeds the number, N, of available places, the topmost N applicants in the scoring will be awarded the places, with some mechanism for breaking a tie of scores).

5.10. **Applicant Preferences.** It is assumed that applicants will continue to specify their preferences for foundation schools, and it will be the applicants' preferences, combined with their national score, that will determine who is allocated to which school. The actual choice of matching algorithm is one of the 'secondary' options.

5.11. **Eligibility.** It is assumed that GMC will continue to be the authority for determining the professional eligibility of applicants.

5.12. **Special Cases.** It is assumed that, as with the current system, separate consideration will be given to 'special circumstances' applicants, academic posts, and military trainees.

5.13. **National Portal.** It is assumed that there will continue to be some national system through which: applicants will submit their details and preferences; foundation schools will submit details of

[3] Almost all students from UK medical schools apply for the Foundation Programme, so the integrity of the quartiles can be simply assured by checking that each quartile contains a quarter of the applicants from a school. Such assurance cannot be obtained where a small number of students apply from a school.

their posts; scores will be entered; and applicants matched to schools through an appropriate algorithm.

5.14. **National Timetable.** It is assumed that the recruitment activity will be coordinated, where necessary, to a common timetable, with broadly the same timing and duration as the current process (see Section 3).

5.15. **Matching to posts.** It is assumed that the scores used to match applicants to Foundation Schools will remain available to the schools to be used for matching applicants to posts within the school.

5.16. **Central administration.** It is assumed that there will continue to be a need for some central administration to undertake the work currently done by UKFPO which is common to all the options, such as: gaining agreement for the timetable for a recruitment round; checking the eligibility of applicants from non-UK medical schools; publishing guidance for applicants and those involved in processing the applications; quality assuring the application of the selection method; dealing with FOI requests; managing the National Portal, and so on.

5.17. The options are considered in turn below.

Do Nothing

5.18. This option is to retain the current recruitment process broadly as it is. For the purpose of the analysis, the key facts about the current approach (over and above those summarised in Section 3) are as follows:

5.19. The answers to the white space questions are scored by volunteers from across the NHS, who participate in scoring and calibration workshops to encourage a standard approach.

5.20. The scoring is organised by the foundation schools, with each school being responsible for those applicants who have put the school as their first choice; this is meant to reduce (or at least to standardise) any bias in the marking. The scorers cannot see the identities of the applicants they are scoring; this is intended to reduce bias. The scoring is performed 'horizontally' (i.e. a group of scorers asses the answers to a specific question from multiple applicants, rather than all of the answers from a single applicant); this is intended to reduce the impact of possible 'hawks and doves' among the markers. Random samples of applications are re-marked to check the calibration of the scoring and to determine the incidence of plagiarism.

5.21. The white space questions are newly drafted each year by representatives of the Foundation Schools, and piloted with groups of F1 doctors. The questions are reflective in nature, and designed to test whether the applicant has the qualities set out in the national person specification for F1 posts.

Multiple Mini Interviews

5.22. In this option multiple, short interviews replace the use of white space questions to score applicants. The interviews would be designed to assess the broader professional attributes of the applicants linked to the personal specification for foundation training, and the results would be combined with the educational performance scores provided by the medical schools.

5.23. The term 'Multiple Mini Interviews' (MMI) describes an arrangement in which each person being assessed undergoes a consecutive set of short interviews, each designed to assess a particular subset of the qualities being investigated.

5.24. For the purpose of costing and assessing this option, the following assumptions apply:

5.25. The Multiple Mini Interviews would be undertaken by every applicant. There would be six interviews in the cycle, each lasting ten minutes with one interviewer present. Three members of each interview team would be clinicians; two would be senior management staff; one would be a trained lay person. Each interview team would spend half a day at the start of the process preparing for the interviews, and half a day at the end on a wash-up. New interviewers would need to be trained in good interview practice- for the purpose of the costings it is assumed that half of the interview panel members attend half a day of such training.

5.26. The arrangements for the interviews would be the joint responsibility of the Foundation and Medical Schools. These arrangements will include: advertising details to applicants; making and revising bookings; collating interview notes and scores; entering scoring information to be used for allocation.

5.27. The requirements for defining and quality assuring the interviews (including the design of questions, calibration, piloting, sampling for consistency and so on) would have broadly the same resource implications as the corresponding activities undertaken currently in relation to white space questions.

5.28. UK applicants will attend interviewed near their local medical school. The interviews would be timed to avoid any clash with electives.

5.29. Overseas applicants will attend interviews at their first choice school.

Structured Interviews

5.30. This option is identical to MMI, except that each applicant undergoes a single interview with a panel of interviewers, rather than a series of short interviews. Again, the interview would be designed to assess the broader professional attributes of the applicants linked to the personal specification for foundation training , and the results would be combined with an educational performance score provided by the medical schools. For the purpose of costing it is assume that:

5.31. - the interview panel will include three members, two clinicians and one trained lay representative.

5.32. - each panel will complete 12 interviews per day.

5.33. -each panel will spend half a day at the start of the process preparing for the interviews, and half a day at the end on a wash-up.

5.34. -new interviewers would need to be trained in good interview practice- for the purpose of the costings it is assumed that half of the interview panel members attend half a day of such training.

National Assessment for Ranking

5.35. In this option all applicants would sit a standard national assessment of their clinical knowledge and skills, the scores from which would be used for selection and allocation.

5.36. For the purpose of costing and assessing this option, the following assumptions apply:

5.37. The assessment would be undertaken by all applicants under invigilated conditions.

5.38. The assessment would be held on a single day, so that questions from earlier sittings could not be leaked to those sitting in later ones. The sensitivity analysis will consider the impact of holding more than one assessment (with test-equating) to accommodate students away on their elective.

5.39. The assessment would be a combination of machine-markable knowledge tests, and an OSCE.

5.40. The arrangements for the tests would be the joint responsibility of the UK Foundation and Medical Schools. These arrangements will include: advertising details to applicants; confirming attendance; invigilation, organising and training the scoring panels, collating scripts and scores; entering scoring information to be used for allocation.

5.41. The steps to define and quality assuring the assessment (including setting the questions, piloting them, sampling for consistency and so on) would be broadly the same as those to achieve the corresponding objectives in relation to white space questions; however, the resource required will be greater (see section 9).

5.42. When taking into account the cost of travelling to and from the examination venues, it will be assumed that around 10% of UK students would need to return from electives abroad if there was only one assessment day (see 5.36).

Educational Performance

5.43. In this option there would be no separate test or assessment for the purpose of selection. Instead the Medical Schools would provide scores for their students based on a range of academic and clinical criteria that are comparable across the schools.

5.44. While the label ‘Educational Performance’ is a new one, invented for the purpose of this report, the option itself is a major extension of academic quartiles that has been the subject of consideration by the Expert Panel and the FPSG, namely the wider use of information about applicants which is available within the medical schools. There would be a common national framework as to how these scores are to be derived by schools and transparency on how these have been put into operation.

5.45. For the purpose of costing and assessing this option, the following assumptions apply:

5.46. The activities required to set up and coordinate a common approach across medical schools will be broadly comparable to the ‘Rules Group’ costs for the ‘White Space’ option. The costs of agreeing and piloting the use of the common approach will the broadly the same as the corresponding costs for the other options.

5.47. The administrative effort required to compile the scores would be broadly comparable to that required presently to produce academic quartiles.

5.48. The proposals that have been developed by the Medical Schools Council Assessment Alliance for the use of comment content of assessments will be implemented. This will provide additional assurance of the comparability of assessments across medical schools.

Situational Judgement Test

5.49. The Expert Panel has identified that 'Situational Judgement Tests' (SJT) could be an effective way of assessing the broader professional attributes of prospective doctors in a way that could bolster the validity of the selection process. These tests present the applicant with a variety of hypothetical professional situations, and ask the applicant to chose, from a list, what they consider to be the most appropriate response to that situation.

5.50. For the purpose of costing and assessing this option, the following assumptions have been made:

5.51. The test would be undertaken by all applicants under invigilated conditions, and would be designed to assess their professional judgement and likely behaviours in scenarios based on complex situations encountered in Foundation training.

5.52. Applicants would record their answers on paper, the answers being markable by machine.

5.53. The arrangements for the tests would be the joint responsibility of the UK Foundation and Medical Schools. These arrangements will include: advertising details to applicants; confirming attendance; invigilation, collating test scripts and scores; entering scoring information to be used for allocation.

5.54. The SJT score would be combined with an educational performance score provided by medical schools.

Matching Rules

5.55. Successful applicants to the Foundation Programme are allocated to individual foundation schools by following rules that take into account the applicants' scores and preferences, and the availability of spaces at the schools. There is no absolutely right or wrong way to design such rules- which are known as matching algorithms- although there are some general measures which can be used to classify rules in terms of their effectiveness. The Expert Panel has considered the effectiveness of the current matching algorithm in connection with two of these measures particularly.

5.56. The first is the extent to which higher scoring applicants are matched to places ahead of lower scoring applicants. The algorithm currently used puts applicants with higher scores ahead of those with lower scores in all cases bar one. The exception is that where two applicants are competing for the last place at a school, a lower-scoring applicant for whom the school is a first choice will be given the place ahead of a higher scoring applicant for whom the school is not a first choice. This exception was built-in because it maximises the number of applicants who get their first choice school, this being seen as a desirable outcome from an educational perspective. The effect is not great statistically. Roughly speaking, based on the profiles of the F1 recruitment rounds to date, a few hundred more applicants (out of c7000) can be expected to get their first choice as a result of the exception.

5.57. The second measure is the extent to which the rules encourage the applicants to declare their genuine preferences. Consider the hypothetical case in which an applicant's real first preference is a popular school for which there will be more applicants than places. With the current algorithm, because of the exception described above, a weaker applicant might well decide that they would be better off specifying some other school as their first choice. If the exception was removed, then the applicant would have nothing to lose by declaring their 'true' preferences.

6. Assessment Criteria

6.1. As described in Section 4, best practice calls for each of the options to be scored against a weighted set of assessment criteria. The criteria need to be carefully defined so that they focus on the impacts of the options, rather than any other characteristics the options might have, the aim being to weigh up the beneficial and adverse impacts of the options in a systematic way. An ideal set of criteria would be:

Complete. They would cover all of the ways in which the impacts of the options might differ.

Unambiguous. The criteria should be clear, and not open to interpretation.

Scoreable. It should be possible to score each option in a meaningful way against each of the criteria.

Non-overlapping. There should be no unintended double-counting between scores for different criteria.

6.2. In practice it is not be possible to attain these ideal characteristics completely. Instead a pragmatic approach is to start with a working set of criteria that is broadly correct, and then correct any deficiencies found through their application.

6.3. In the present work the assessment criteria were developed in the following way. The FPSG had developed an informal set of criteria to guide their initial considerations. These criteria were then tested against the ideal characteristics listed above, and a refined set put forward for consideration. A further test for completeness was performed by mapping the criteria to the 450 or so individual observations on the pros and cons of possible options that had been collated by the FPSG through facilitated sessions with stakeholder groups.

6.4. The resulting set of criteria agreed by the FPSG is shown in the table below.

Name	Definition
Reliability	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.
Validity	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.
Granularity	The extent to which the selection method can separate out the performance of comparable applicants for the purpose of scoring.
Consistency	The extent to which the selection techniques can be expected to be consistently applied across the UK.
Longevity	The extent to which the performance of the option can be maintained over successive recruitment rounds.
Educational Impact	The extent to which the option supports or undermines educational objectives.
Fairness	The extent to which the option offers a level playing field for applicants.
Compliance	The extent to which the option discourages, prevents, or otherwise guards against the effects of cheating or malpractice.
Transparency	The ease with which applicants can understand what is expected of them and why they achieved their ranking.
Applicant Burden	The extent to which the option minimises the costs and effort for applicants.
Medical Time	The extent to which the option minimises the amount of medical staff time required for selection.
Feasibility	The ease with which the option could be brought into successful live use.
Public Opinion	The extent to which the general public, when presented with a sound-bite description of the process, are likely to agree with it.

6.5.

6.6. The assessment criteria were then subject to review by stakeholders at consultation events held on 16th June 2009, at which some other criteria were suggested, as follows:

Adaptability- the ease with which changes could be made to the option. This would appear to be an important consideration in absolute terms; however, it does not seem to vary significantly in relation to the options under consideration, so it has been excluded from the analysis.[4]

Administrative time- the effort required to administer the option. The variation in administrative effort across the options is considered to be fully taken into account via the cost models, so administrative time has not been included as a non-monetary assessment criterion. The justification for including such a criterion for clinical time is the assumption that the day rates for clinical staff probably under-estimate the true opportunity cost of distracting skilled clinical staff from service delivery.

Legality- the extent to which the option would withstand legal challenge. This has been excluded from the current analysis for two reasons: firstly, the criteria of Fairness, Consistency, Validity, and so on measure the factors that might determine the legality of the options; secondly, the legality of the options will be the subject of a separate legal review outside the scope of the CBA work.

[4] All of the options have some level of adaptability- such as the scope for introducing new questions, changing scoring and weighting rules, coping with changes to numbers of applicants, and so on. If a more significant change were required, then it is likely that such a change would require investigation and piloting regardless of which option were chosen.

7. Common Costing Factors.

7.1. Some of the costing factors are common to all of the options considered. These common factors are summarised below then considered in more detail.

7.2. **Scope and timescales.** The costs have been modelled over five years, and cover the piloting and introduction of the option, and the running of four subsequent annual F1 recruitment rounds. The costs incurred by Foundation Schools in allocating applicants to individual posts are excluded from the models except insofar as these costs vary by option.

7.3. **Applicants and Posts.** It is assumed that 8000 applications will be processed each year, with 5% of these being from applicants who are not graduating from UK medical schools. The sensitivity analysis will consider the likely impact on the options of significant changes to the number of applicants, and the ratio of applicants to posts.

7.4. **Schools.** It is assumed that the number of Foundation Schools and Medical Schools will remain unchanged over the evaluation period.

7.5. **Year 1 Round.** The piloting and development of any new option would have to be done in parallel with the continuing use of the existing process. Therefore the year 1 costs of the 'White Space' option are included in the NPV calculations for all other options.

7.6. **Central Administration and Allocation System.** All of the options will require some form of central administration and a system to manage the allocations based on scores and preferences. It is assumed that the cost of this will be equal to the actual costs for performing the corresponding functions in relation to the current F1 recruitment process, which was just under £1m for the 2008/9 F1 round.

7.7. **Procurement.** Regardless of the option selected, there will be a requirement to re-procure the central administrative services and the national application system. These costs have been taken to be £200,000 and are assumed to be incurred in 2010/11.

7.8. **Manpower.** Where NHS resource costs have been modelled in person-days, the opportunity costs have been modelled using four representative day rates.

7.9. **Piloting.** All of the options will require piloting and development.

7.10. **Common Content of Assessment.** For all of the options bar National Assessment, a standard cost has been included for the delivery of the Common Content of Assessment initiative.

7.11. **Legal Challenge.** The sensitivity analysis considers the possible impact of legal challenges to the process. For this purpose the CBA will consider a standard unit cost for defending a legal challenge, based on actual costs incurred by DH in comparable circumstances, and factor these into the analysis by considering the likelihood and impact of an option resulting in one or more legal challenges.

7.12. **Net Present Values (NPVs)**. These have been calculated from the cost estimates according to 'Green Book' rules. The NPVs include some allowance for optimism bias[5], but mainly optimism bias is considered through sensitivity analysis.

7.13. These common factors are discussed in more detail below

[5] Optimism bias refers to the tendency for business cases to under-estimate costs.

Scope and Timescales

7.14. The scope of the cost models is deliberately limited to those elements of cost that are directly associated with the variety of options under consideration, namely the costs to run the first stage of a national recruitment round. The upstream and downstream activities- such as the costs of performing pre-employment checks, or of delivering undergraduate training - are excluded from the base cost models, but where they can be expected to vary by option, the variances will be taken into account.

7.15. The annual costs in the models are based on public sector financial years rather than calendar years, as this aligns better with the timing of the annual recruitment rounds.

Applicants and Posts

7.16. The number of applications received each year will include a dominant proportion from UK medical schools and a small remainder from schools outside the UK. The former figure can be predicted with some confidence for several years ahead, as it is broadly the total number of students in the corresponding years at medical school[6]. However, the latter figure could be subject to large changes from one year to the next. The historical trend has been as follows:

Year	Total Applicants	Non-UK applicants
2006	6467	295
2007	6430	412
2008	7004	186

7.17. On this basis, 8000 has been taken to be a sensible ball-park figure for the total number of applicants each year, with 5% assumed to be from schools outside of the UK. The effects of variation of these assumptions will be considered in the sensitivity analysis.

7.18. To date the number of available posts has always exceeded the number of applicants. The sensitivity analysis will consider the impact of changes to the ratio of applicants to posts.

Schools

7.19. The number of medical schools and foundation schools is open to change, through splits and mergers for example. In the current models the possibility of such changes is ignored, on the grounds that the costs are expected to relate to the number of applications being dealt with, and not to the way in which the workload is spread-out between individual schools. The validity of this assumption is confirmed in the sensitivity analysis.

Central Administration and Allocation System

7.20. All of the options will have some form of central administration to cover the following:

- The definition of a standard timetable for the round
- The definition of eligibility criteria
- Maintaining the standard application form
- Development and publication of guidance for applicants and selectors
- Provision and management of the national allocation system (see below)

[6] The numbers cannot be forecast exactly, as there will be students who drop out, or who re-sit years, etc.

- Management reporting to stakeholders
- Registration of eligible applicants
- Checking eligibility of applicants from overseas medical schools
- Publication of the results of allocation
- Quality assurance of the national process
- Dealing with FOI and DPA queries.

7.21. The options also all assume that there will be a central allocation system through which: foundation schools will provide information about the number and type of their available training posts; medical schools will enter details of their students; applicants will submit applications and state their preferences; the scores for the applicants will be entered; a matching algorithm will be used to allocate applicants to schools based on scores, preferences, and the availability of posts.

7.22. The base financial model will assume that for all options these costs will be equal to the actual costs for performing the corresponding function in relation to the current F1 recruitment process, which was just under £1m for the 2008/9 F1 round. The sensitivity analysis will consider the possible impact of these costs varying by option.

Procurement

7.23. The central administration services are currently provided through two main contracts, one covering the provision of the 'Application MMC' national application system, the other covering the administration and management services. Both of these contracts will expire during the evaluation period, so some allowance has to be made for re-procurement. It is assumed that the cost of managing the re-procurement (taken to be £200,000) will not vary by option; this assumption will be revisited in the sensitivity analysis.

Manpower

7.24. The manpower requirements for the selection options can be broken down into four main categories for costing purposes: clinical (e.g. to conduct interviews); senior management (e.g. to define standards and rules); lay members of interview panels, and administrative time (e.g. to arrange interviews, or collate scores).

7.25. There is no authoritative yardstick for measuring the opportunity costs of NHS or medical school staff. Given this, the approach adopted in the current work has been to consider market prices for temporary staff as measure of the opportunity cost. In effect this is saying that whatever an organisation is prepared to pay to backfill a temporarily empty role is a measure of the opportunity cost of leaving it vacant. This principle is most likely to hold true where resources are in short supply, as there the market price is driven up towards a limit which should be the opportunity cost. Given this, the market price is probably is a reasonable measure for the clinical and senior management roles.

7.26. For the current work the following day rates have been adopted, arising from discussions with locum agencies and inspection of random samples of job adverts for temporary positions:

- Clinical £650 per day
- Senior Management £500 per day
- Lay Member £250 per day

- Administrative £150 per day

7.27. The impact of the assumed rates is considered in the sensitivity analysis.

Piloting and Development

7.28. All of the options will require piloting and development- even the 'Do Nothing' option, since the current system would have to be re-procured within the evaluation period. For the purpose of building the base cost models it is assumed that the piloting and development would present broadly similar resource requirements regardless of which option is chosen, and can be represented by an order of magnitude figure of £1.5m. The exception to this is the 'Do Nothing' option, where the scope of the piloting is reduced, for which the pilot cost is taken to be £0.75m. The sensitivity analysis will consider the possible impact of variations to these assumptions.

Common Content of Assessment

7.29. In all of the options bar National Assessment it is assumed that the Common Content of Assessment initiative is implemented across all UK medical schools. The cost models include the estimates for this work that have been provided by the Medical Schools Council, which amount to £1.5m in the first year and £1m per year thereafter.

Legal Challenge

7.30. It can be expected that the risk of legal challenge will vary across the options. The impact of this will be considered in the sensitivity analysis by considering the extent to which the cost of legal challenges could be a factor in determining a preferred option.

7.31. As an indication, the DH has confirmed that the legal costs of two judicial reviews in recent years have totalled around £150,000. Given that these exclude DH management time, it is proposed here that the cost of successfully defending a single challenge should be taken to be £100,000.

7.32. The impact of losing a legal challenge is far more difficult to assess, as the range of possible outcomes is very broad. A worst case might be that a prevailing selection process has to be abandoned and replaced by another at short notice, possible with some interim working arrangements. This will be modelled in the present work as an impact of £2.1m, which is built up as follows:

- Cost of defending the challenge: £150,000
- Cost of consultation and design of new option £250,000
- Cost of piloting and development of new option £1,500,000
- Cost of procuring new option £200,000

NPV Calculations

7.33. The NPV calculations are based on 2008/9 figures, discounted annually by 3.5%. VAT is excluded from all cost estimates.

7.34. Given the nature of the options under consideration, there are no significant or systematic distributional effects to take into account, except that some options shift the burden of assessment from employers and Foundation Schools to Medical Schools.

7.35. Treasury and DH guidance is that the evaluation of options should take into account an effect called 'optimism bias', which is the tendency for business cases to underestimate costs and over-estimate benefits. For some types of capital spend, where there is good historical evidence

from projects, the Treasury specifies a percentage uplift that should be added to costs to take account of optimism bias. For other types of capital spend, where there is no established evidence, and for operational costs and benefits, the Treasury guidance is that the effects of optimism bias should be considered through sensitivity analysis. In the current work, for a small proportion of the overall costs- namely the estimated £200k of costs included in each year for making changes to the central application system- there is a relevant uplift factor (200%) provided by Treasury which can be considered relevant. This factor has therefore been added to the NPVs of all of the options, allowing an extra £400k in each of years 2-5 for uncertainties in the exact specification of the central application system. Other effects of optimism bias are considered through the sensitivity analysis.

8. Monetary Analysis

8.1. This section considers the relative values of each option in terms of those costs and benefits that can be quantified in monetary terms. It considers each of the options in turn, and ends with a summary.

Do Nothing

8.2. The table below summarises the monetary cost elements for the Do Nothing option, over and above those that are common costs, as defined in Section 7.

Name	Description	Basis of Estimate	Annual Estimate (£,000.00)
Rules	The costs of defining the questions, and assuring their effectiveness.	Figures from UKFPO.	225
Educational scores and registration	The administrative costs incurred by medical schools in providing applicant details to be loaded onto the system.	See below.	100
Scoring	The non-central manpower costs incurred in scoring applications.	Manpower figures from UKFPO.	1558
Venues	The non-manpower costs associated with scoring applications, including room-hire, catering, travel expenses etc.	See below.	125

8.3. The Rules costs are those associated with the following activities:

- Drafting the white space questions and associated scoring guidelines, through consideration of the person specification
- Validating the questions and guidance through piloting and consultation.
- Holding calibration workshops
- Reviewing special cases (e.g. examples of suspected plagiarism).

8.4. The UKFPO has estimated these costs at 450 person days, the bulk of which is senior management time. For the purpose of the current estimates it is assumed that all the time can be rounded up to the senior management manpower rate, and the cost of travel, room hire and so rounded down to zero, giving a cost estimate of £225,000.

8.5. The educational score and registration costs are those incurred by the medical schools in providing scores for all of their applicants, and presenting those scores, plus the email addresses of their students, in an electronic form to be loaded on the national system. This cost has been estimated on the basis of the following breakdown of activities per school.

Activity	Admin days per school	Management days per school	
Familiarisation with rules for current year	2	2	
Identify sources of information	2	1	
Download email addresses of all current students to file	1		
Collate scoring information	2		
Approve quartiles	0.5	0.5	
Upload information to central system	1		
Contingency	1	1	
Totals	9.5	4.5	
Cost per school	£1,425	£2,250	£3,675

8.6. The total for all medical schools has been rounded to £100,000[7].

8.7. The scoring costs cover the manpower required to organise scoring events, train scorers, print the online forms, score the forms, collate the scores and enter them onto the system.(In the sensitivity analysis there will be consideration of a variant in which the application forms are scored online, which is an option supported by the current system.) These costs have been estimated as follows (based on UKFPO figures for 7,000 applications), assuming that half of the scorers are clinicians and half are senior managers:

Activity	Admin Days 7,000 Applications	Senior/ Clinical Days 7,000 Applications	Cost 7,000 Applications £000	Cost 8,000 Applications £000
Organise events and print application forms	910		137	156
Score applications		2000	1150	1314
Manage scoring events	65	65	42	48
Collate and enter scores	230		35	39
Total	1205	2065	1363	1558

8.8. The venue costs have been estimated at £5,000 per school, based on informal figures provided by foundation school representatives at a meeting of the Foundation Programme Rules Group in 2008.

[7] This figure may be pessimistic. Some medical schools have reported that score information can be straightforwardly downloaded from their administrative systems with little manual effort.

8.9. Given the above, an illustrative NPV for the Do Nothing option of -£21.2 million can be calculated as shown below:

Costs (£000)		Year 1	Year 2	Year 3	Year 4	Year 5	Total	
Central Administration		1,000	1,000	1,000	1,000	1,000	5,000	
Common Content			1,500	1,000	1,000	1,000	4,500	
Rules		225	225	225	225	225	1,125	
Registration and EP Scores		100	100	100	100	100	500	
Scoring		1,558	1,558	1,558	1,558	1,558	7,790	
Venues		125	125	125	125	125	625	
Procurement			200				200	
Pilot		0	0	1,500	0	0	1,500	
Optimism bias factor			400	400	400	400	1,600	
Total		3,008	5,108	5,908	4,408	4,408	22,840	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		3,008	4,929	5,502	3,961	3,823	**21,223**	**NPV**

Multiple Mini Interviews

8.10. The table below summarises the monetary cost elements for the Multiple Mini Interviews option, over and above those that are common costs, as defined in Section 7.

Name	Description	Basis of Estimate	Annual Estimate (£,000.00)
Rules	The costs of defining the interview questions and guidance, and assuring their effectiveness.	Assumed to be equivalent to the Do Nothing costs	225
Registration and EP scores	The administrative costs incurred by medical schools in providing applicant details to be loaded onto the system.	As per 'Do Nothing' costing.	100
Interviewer time	The manpower costs of the interviewers.	First-principles estimate, detailed below	2600
Interview Admin	The non-central manpower costs or arranging and preparing for interviews, managing them, and collating information about outcomes.	As per 'Do Nothing' event admin, as described below.	244
Venues	The non-manpower costs associated with scoring applications, including room-hire, catering, travel expenses etc.	See below.	368

8.11. The manpower estimates for the interview panels have been developed as shown below, based on the assumptions presented in Section 5.

Interview cycles required:	8000	
Each team does:	15	cycles per day
Each team interviews for:	3	days
So each team does:	45	interview cycles
So the number of teams needed is:	177.77778	
		day on prep and a wash-
The team also spends	1	up
So each team works for:	4	days
The total number of team days is:	711.11111	

Assume 3 clinical, 1 lay, and 2 senior members per team.	
Average cost per team day is:	£3,200
So the cost of interviewing time is:	£2,275,556
The cost of training half the teams is:	£142,222
So the total cost is:	£2,417,778

8.12. The resources required to prepare for and manage the interview events are taken to be broadly comparable to those required for the scoring events for the 'Do Nothing' option. The rationale for this is that the steps involved are quite similar. The two main differences are:

8.13. -There will be no requirement to print out lengthy application form answers in the case of MMI. That said, there will still be a requirement to print-out interview sheets and schedules.

8.14. -There is no requirement to book applicants in to events in the case of 'Do Nothing', and the scoring events will be shorter overall than the interview events.

8.15. It is reasonable to assume that these two differences will broadly cancel out in cost terms.

8.16. As with the 'Do Nothing' option, where the marks for individual answers are entered onto the national system, it is expected that the scores from each individual interview in the cycle will need to be recorded for audit purposes.

8.17. Given this, the admin costs are as per 'Do Nothing' less the costs of the scoring, namely:

Activity	Admin Days for 7,000 Applications	Senior Days for 7,000 Applications	Cost for 7,000 Applications £000	Cost for 8,000 Applications £000
Organise events and print application forms	910		137	156
Manage scoring events	65	65	42	48
Collate and enter scores	230		35	39
Total	1205	65	213	244

8.18. At a meeting of the F1 Rules Group in 2008, representatives of foundation schools indicated that the cost of room hire and catering for hosting interviews would be of the order of £10,000 per school, or £250,000 for the UK.

8.19. Assuming that interviews are held at venues local to the UK applicant, a figure of £5 per applicant has been assumed for UK travel expenses, giving £38,000 for 7,600 applicants. Costs for a single non-UK applicant have been assumed to be £200, giving £80,000 for 400 applicants.

8.20. From the above, total Venue costs (including applicant expenses) are £368,000.

8.21. Given the above, an illustrative NPV for the MMI option of -£26.3 million can be calculated as shown below:

Costs (£000)		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Total**	
Central Administration			1,000	1,000	1,000	1,000	4,000	
Common Content			1,500	1,000	1,000	1,000	4,500	
Rules			225	225	225	225	900	
Registration			100	100	100	100	400	
Interview Teams			2,418	2,418	2,418	2,418	9,671	
Interview Admin			244	244	244	244	975	
Venues			368	368	368	368	1,472	
Procurement			200				200	
Pilot		1,500					1,500	
Year 1 round		3,008					3,008	
Optimism bias factor			400	400	400	400	1,600	
Total		4,508	6,454	5,754	5,754	5,754	26,626	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		4,508	6,229	5,359	5,171	4,990	**26,257**	**NPV**

Structured Interviews

8.22. The annual costs of interviewer time for structured interviews can be estimated as follows:

Interviews required	8000	
Each panel manages	12	interviews per day
Each panel interviews for	3	days
So each panel does	36	interviews
So the number of panels needed is	222.22222	
Each panel also spends	1	day on prep and a wash-up
So each panel works for	4	days
So the total number of Panel Days is	889	

Assume 2 clinical and 1 lay member per panel.	
Average cost per panel day	£1,550
Cost of training half the panels	£86,111
Total cost	£1,463,889

8.23. In other respects, the costs for the Structured Interviews option will be comparable with those for MMI (a possible exception- namely the cost of venues- is considered in the sensitivity analysis).

Given this, an illustrative NPV for the Structured Interview option of -£22.8 million can be calculated as shown below:

Costs (£000)		Year 1	Year 2	Year 3	Year 4	Year 5	Total	
Central Administration			1,000	1,000	1,000	1,000	4,000	
CCA			1,500	1,000	1,000	1,000	4,500	
Rules			225	225	225	225	900	
Registration			100	100	100	100	400	
Interview panels			1,464	1,464	1,464	1,464	5,856	
Interview Admin			244	244	244	244	975	
Venues			368	368	368	368	1,472	
Procurement			200				200	
Pilot		1,500					1,500	
Year 1 round		3,008					3,008	
Optimism bias factor			400	400	400	400	1,600	
Total		4,508	5,501	4,801	4,801	4,801	24,410	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		4,508	5,308	4,470	4,314	4,163	**22,763**	**NPV**

National Assessment for Ranking

8.24. The table below summarises the monetary cost elements for the National Assessment for Ranking option, over and above those that are common costs, as defined in Section 7.

Name	Description	Basis of Estimate	Annual Estimate (£,000.00)
Registration	The administrative costs incurred by medical schools in providing applicant details to be loaded onto the system.	As per 'Do Nothing' costing.	100
Examinations	The all-in costs of developing and delivering the test.	GMC PLAB costs (see below),	4600
Expenses	The cost of applicant travel expenses.		274

8.25. The estimates for the examinations are based on the charges imposed by the GMC for the PLAB examinations, which are £145 and £430 per student for Parts 1 and 2 of the test respectively. These charges are designed to recover, without profit, the GMC's costs for the development, maintenance and delivery of the examinations.

8.26. It is assumed that at least 10% of UK students would be away on electives whenever the examination was held, and would have to return to the UK for the examination. Treating these applicants in the same way as the 5% of overseas applicants, the cost for expenses is £274,000. The impact of this assumption will be considered in the sensitivity analysis.

8.27. Given the above, an illustrative NPV for the National Assessment for Ranking option of -£32.2 million can be calculated as shown below:

Costs (£000)		Year 1	Year 2	Year 3	Year 4	Year 5	Total	
Central Administration		0	1,000	1,000	1,000	1,000	4,000	
Common Content		0	1,500	1,000	1,000	1,000	4,500	
Rules		0	0	0	0	0	0	
Registration		0	100	100	100	100	400	
Examination		0	4,600	4,600	4,600	4,600	18,400	
Venues (expenses only)		0	274	274	274	274	1,096	
Procurement			200				200	
Pilot		1,500					1,500	
Year 1 round		3,008					3,008	
Optimism bias factor			400	400	400	400	1,600	
Total		4,508	8,074	7,374	7,374	7,374	34,704	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		4,508	7,791	6,867	6,627	6,395	**32,187**	**NPV**

Situational Judgement Test

8.28. The table below summarises the monetary cost elements for the SJT option, over and above those that are common costs, as defined in Section 7.

Name	Description	Basis of Estimate	Annual Estimate (£,000.00)
Registration	The administrative costs incurred by medical schools in providing applicant details to be loaded onto the system.	As per 'Do Nothing' costing.	100
Examinations	The all-in costs of developing and delivering the tests.	GMC PLAB costs (see below),	1160
Rules	The costs of determining the rules that relate to the use of an educational performance score.	See below.	115
Expenses	The cost of applicant travel expenses.	As per MMI	80

8.29. The estimates for the examinations are based on the charge imposed by the GMC for Part 1 of the PLAB examinations, which is £145 per student. These charges are designed to recover, without profit, the GMC's costs for the development, maintenance and delivery of the examinations.

8.30. The 'Rules' costs here are roughly estimated as half the corresponding costs for other options such as MMI. The rationale is that for those other options the Rules costs cover two areas: 1) the use of an assessment technique, such as MMI; and 2) the use of some educational performance score. With SJT , the first of these things- namely the cost of defining the SJT itself- is modelled by the all-in PLAB costs, so only the costs associated with the use of some Educational Performance score need to be included. The effects of this estimate are considered in the sensitivity analysis.

8.31. Given the above, an illustrative NPV for the SJT option of -£19.3 million can be calculated as shown below:

Costs (£000)		Year 1	Year 2	Year 3	Year 4	Year 5	Total	
Central Administration			1,000	1,000	1,000	1,000	4,000	
Common Content			1,500	1,000	1,000	1,000	4,500	
Rules			115	115	115	115	460	
Registration			100	100	100	100	400	
Examination			1,160	1,160	1,160	1,160	4,640	
Venues (expenses only)			80	80	80	80	320	
Procurement			200				200	
Pilot		1,500					1,500	
Year 1 round		3,008					3,008	
Optimism bias factor			400	400	400	400	1,600	
Total		4,508	4,555	3,855	3,855	3,855	20,628	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		4,508	4,396	3,590	3,464	3,343	**19,301**	**NPV**

Educational Performance

8.32. The table below summarises the monetary cost elements for the Educational Performance option, over and above those that are common costs, as defined in Section 7.

Name	Description	Basis of Estimate	Annual Estimate (£,000.00)
Registration and scores	The administrative costs incurred by medical schools in providing applicant scores and details to be loaded onto the system.	As per 'Do Nothing' costing.	100
UK Rules	The costs of defining and maintaining the rules for generating medical school scores	As per 'Do Nothing' costing	225

8.33. Note that the costs incurred by overseas medical schools in providing scores for their applicants are excluded from the calculations because they are not incurred within the UK.

8.34. There is an argument to say that the Rules costs for the EP option should be less than for the other options (e.g. MMI) in which some educational performance score is combined with another assessment technique, the argument being that those options need rules for the assessment technique as well as for the use of the educational score. However, there is a counter argument which is that where an educational performance score is the only measure being used for selection, it would need to be a more robust and granular score, hence the associated rules need to be considered in more detail. Given this, the two competing effects represented by these arguments have been taken to cancel- out; the sensitivity analysis examines the importance of this assumption.

8.35. Given the above, an illustrative NPV for the Educational Performance option of -£15.2 million can be calculated as shown below:

Costs (£000)		Year 1	Year 2	Year 3	Year 4	Year 5	Total	
Central Administration			1,000	1,000	1,000	1,000	4,000	
Common Content			1,500	1,000	1,000	1,000	4,500	
Registration and scores			100	100	100	100	400	
Rules			225	225	225	225	900	
Procurement			200				200	
Pilot		1,500					1,500	
Year 1 round		3,008					3,008	
Optimism bias factor			400	400	400	400	1,600	
Total		4,508	3,425	2,725	2,725	2,725	16,108	
Discount Factor	3.50%	1.00	0.97	0.93	0.90	0.87		
Discounted Costs		4,508	3,305	2,538	2,449	2,363	**15,163**	**NPV**

Summary

8.36. The table below summarises the NPV calculations for the options, based on the assumptions made in this section.

Option	NPV (£000)
Do Nothing	-21,223
MMI	-26,257
National Assessment	-32,187
Educational Performance	-15,163
Structured Interview	-22,763
SJT	-19,301

8.37. These can also be expressed relative to the NPV of the 'Do Nothing' option, as follows:

Option	Relative NPV (£000)
Do Nothing	0
MMI	-5,034
National Assessment	-10,965
Educational Performance	6,060
Structured Interview	-1,541
SJT	1,922

9. Non-Monetary Analysis.

9.1. In this section costs and benefits that cannot be expressed in monetary terms are considered. The section begins by comparing the options against each of the assessment criteria in turn, and explaining the basis upon which scores have been allocated. The weightings that have

been agreed by the FPSG are then introduced, with an explanation of the way in which they have been derived. The section ends with a summary inter-comparison of the weighted performance of the options.

9.2. Before considering the scores it is important to remember that the scoring method used is one known as 'local' scoring (as opposed to 'global' scoring), in which the gap between the best and worst option is always set at 10 points, no matter how significant that gap is in absolute terms. At a later stage a weighting will be introduced to take into account the significance.

Reliability

9.3. Reliability in our terms is a measure of the extent to which a selection method will yield consistent results when properly applied. The scores for reliability are presented in the table below.

Option	Reliability
Do Nothing	0
Multiple Mini Interviews	10
Structured Interview	8
National Assessment for Ranking	10
SJT	10
Educational Performance	8

9.4. The rationale for the scoring is as follows.

9.5. The view of the Expert Panel is that National Assessment, MMI and SJT can be expected to have good reliability, whereas Do Nothing cannot. The main issue with Do Nothing in this regard is that applicants do not complete their answers under invigilated conditions, so that some applicants might be able to spend far more time on their answers than others (leaving aside the possibility of cheating, which is dealt with under a separate criterion below).

9.6. The Panel regarded Structured Interview as marginally less reliable than MMI.

9.7. The reliability of Educational Performance ought to be good, on the assumption that the scores awarded to applicants will be based on multiple data points linked to their performance in written examinations, OSCEs and so on, that should themselves be reliable tests.

Validity

9.8. Validity can be considered here as the extent to which selection method is actually measuring the properties that make one applicant 'better' than another. The scores for Validity are presented in the table below.

Option	Validity
Do Nothing	3
Multiple Mini Interviews	8
Structured Interview	7
National Assessment for Ranking	0
SJT	10
Educational Performance	6

9.9. The rationale for the scoring is as follows.

9.10. The Expert Panel reports that there is little evidence about the predictive validity of the proposed methods in comparable applications. This is justified by the output of the literature review, which shows that a lot of work has been done on the assessment of validity, but the results and nature of the investigations vary so widely that it is not possible to draw clear conclusions from them for the purpose of this analysis.

9.11. Given this, one approach to the scoring is to base it upon the view of the Expert Panel that a mix of measures, covering both cognitive and non-cognitive capabilities, will in general have more predictive capability than a single measure. On this basis National Assessment for Ranking is the weakest option, as it is a single measure of cognitive ability only.

9.12. SJT is given the highest score, as it combines an assessment of professional attributes that has good validity, with the academic and clinical components of the educational performance score. It therefore out-points the Educational Performance option, in which the professional attributes are not subject to such a detailed assessment.

9.13. On the same basis, MMI also out-points Educational Performance, although it is not scored as highly as SJT.

9.14. On the assumption that the professional attribute component of the Educational Performance score are based on unstructured interviews, it is marked below the Structured Interview option.

9.15. Do Nothing is marked lower still, as it does not test spoken communication skills, and is not performed under invigilated conditions.

Granularity

9.16. Granularity in this context is a measure of the smallest percentage difference in score that can reasonably be given to comparable applicants. For example, the academic quartile arrangement described earlier in this report can only give one of four scores to any applicant. The less granular the system, the less able is it to distinguish between applicants who are near the border between selection and non-selection. The scores for Granularity are presented in the table below.

Option	Granularity
Do Nothing	5
Multiple Mini Interviews	5
Structured Interview	0
National Assessment for Ranking	10
SJT	10
Educational Performance	8

9.17. The rationale for the scoring is as follows.

9.18. For the Do Nothing option the granularity is governed by the number of questions and the number of possible scores per question. In the latest recruitment round the applicants were graded into 60 categories by score. This could be increased at the cost of increasing the number of questions to be assessed; but that would introduce extra costs for scoring and an extra burden on applicants, assuming that further valid questions could be defined.

9.19. The granularity for MMI is assessed here as broadly the same as 'Do Nothing'. This is on the basis that: the number of interview stations will be broadly the same as the number of scoreable white space questions currently; and that the range of possible marks that could be given to an applicant's performance in each interview station would be broadly the same as the range that could be given to a typical white space question.

9.20. Structured Interview is given the lowest score, as it is unlikely that a single interview of 30 minutes duration could distinguish between applicants to the same extent as any of the other options.

9.21. The National Assessment and SJT have scope to provide high levels of granularity, assuming that the assessments contain large numbers of machine-markable questions.

9.22. Medical Schools are assumed to have sufficient information about their students to place them in a rank order (effectively that is how some schools appear to generate their quartiles at present) which is the highest practical level of granularity achievable in a single school. However, it is not clear whether the common elements of performance reporting that are envisaged in the Educational Performance option would offer such a high level of granularity. Given this, the option has been marked midway between National Assessment and MMI; the sensitivity analysis will determine whether this uncertainty in granularity is an important factor.

Consistency

9.23. Consistency in our terms is the likelihood that a given selection method will be implemented in the same way across the NHS[8]. The scores for Consistency are presented in the table below.

Option	Consistency
Do Nothing	0
Multiple Mini Interviews	2
Structured Interviews	2
National Assessment for Ranking	8
SJT	10
Educational Performance	5

9.24. The rationale for the scoring is as follows.

9.25. The Expert Panel does not comment explicitly on the Consistency of the options, and there appears to be no direct evidence in the output of the literature reviews. Given this, the scores have been arrived at by considering the degrees of freedom that could introduce local variability.

9.26. SJT has the fewest degrees of freedom, given that the tests are standard and are machine markable.

[8] The distinction between Reliability and Consistency in this document is that the former is taken to be the reliability of the method when applied in a given way, and the latter is the likely consistency in the way in which the method is applied across the NHS.

9.27. National Assessment has more degrees of freedom. While some of the assessment is expected to be machine markable, as per SJT, one element is expected to be an OSCE, so there is scope for the OSCE to be organised and run in slightly different ways across the country.

9.28. Both of the interview-based options have more degrees of freedom still, since individuals may also vary in their interpretation of the guidelines for conducting the interviews (in addition to those for scoring).

9.29. Do Nothing has yet more degrees of freedom, as each medical school currently determines its own method for compiling the academic quartiles.

9.30. It is difficult to assess the consistency with which the Educational Performance option could be implemented across the country, and even more difficult to assess how consistently the scores would be awarded by schools outside of the UK.[9] It involves fewer additional steps than the interviewing options and Do Nothing, but more than SJT, say. Given this it is provisionally scored as midway between the two, but this will be re-considered in the sensitivity analysis.

Longevity

9.31. Longevity is the extent to which the quality of the outcome of the process is likely to be sustained over time. The scores for Longevity are presented in the table below.

Option	Longevity
Do Nothing	0
Multiple Mini Interviews	8
Structured Interviews	8
National Assessment for Ranking	9
SJT	9
Educational Performance	10

9.32. The rationale for the scoring is as follows:

9.33. Do Nothing is the only option which allows applicants the freedom to find, and base their answers upon, model answers. What's more, it is possible for those who develop the model answers to get access to the current year's questions and to develop answers to them within the application period. The impact of model answers could grow over time. There would be an increasing likelihood from year to year that the most recent questions would overlap with previous ones. Also, if the competition for places were to grow, the motivation for using model answers would increase. Given these factors, Do Nothing appears to be the worst option by some margin.

9.34. Multiple Mini Interviews and Structured Interviews are marked down slightly because the range of questions that can be put in an interview is reasonably constrained. The National Assessment and SJT, by contrast, would by their nature allow a wider range of questions to be deployed.

9.35. Educational Performance seems to be less open to these difficulties overall, and is thus given the highest score.

Educational Impact

9.36. Educational Impact in our terms is the extent to which the option supports or undermines educational objectives. The scores for Educational Impact are presented in the table below.

[9]For the purpose of selection, the concern relates chiefly to medical schools in the EEA, as applicants from outside of the EEA are only considered for selection if there are posts remaining after UK and EEA applicants have been selected.

Option	Educational Impact
Do Nothing	5
Multiple Mini Interviews	5
Structured Interviews	5
National Assessment for Ranking	0
SJT	5
Educational Performance	10

9.37. The rationale for the scoring is as follows:

9.38. There is broad agreement among medical schools that a national test designed solely for allocation purposes would result in medical schools and students focusing on what it takes to pass the test, rather than focusing on what it takes to develop the qualities needed by a good doctor. An argument could be made that the test should be designed to test the qualities needed to become a good doctor, so the objectives of the test and the objectives of education would be aligned. This argument misses two points. Firstly, while the medical schools deliver a consistently high standard of graduate output, there is considerable diversity across the schools in the way in which that output is achieved, both in terms of teaching styles and the sequencing of the curricula. It would be difficult to define an examination that would be equally fair for all applicants to take several months before finals given that different schools will have completed different parts of the overall learning and development process at that stage. Secondly, even if the exam were aligned with the curricula of the schools, from the applicant perspective the preparation and undertaking of the exam would be distracting from their finals, while overlapping the purpose of their finals. On these grounds the National Assessment for Ranking earns a score far removed from the other options.

9.39. The best option in this regard is Educational Performance, in that it harnesses the efforts that are put into the development of graduates, rather than distracting from it. The remaining options are not well differentiated from each other in relation to this criterion, and are marked down from the best because they all distract the applicant from their studies to a similar extent.

Fairness

9.40. Fairness measures the extent to which all applicants get to play on a level field if a given option is selected. The scores for fairness are presented in the table below.

Option	Fairness
Do Nothing	7
Multiple Mini Interviews	4
Structured Interviews	4
National Assessment for Ranking	0
SJT	10
Educational Performance	5

9.41. The rationale for the scoring is as follows.

9.42. By contrast to the other options, a National Assessment for Ranking requires all applicants to be present in the UK at a certain time, which would present difficulties to those applicants who would be away on electives at the time (there being no suitable time at which no applicants are ever away at electives). Also, it is likely that because of the differences in the curricula across medical schools, it may be that at the time of the assessment some applicants might have very recently covered the topics being assessed, while other will not. For these reasons it has been given the lowest score.

9.43. SJT appears to be a fair option, in that it is anonymised, nationally standardised, should be independent of any variations in curriculum, and gives every applicant an equal amount of time to perform the test.

9.44. Do Nothing could be described as marginally less fair, as some applicants might have more time free within the application period than others.

9.45. Do Nothing scores ahead of the interview options because it is scored anonymously.

9.46. The fairness of Educational Performance is difficult to judge at this stage, so it has been given an average mark, the effects of which will be considered in the sensitivity analysis.

9.47. Note that arguments along the lines of 'exams are unfair because they favour those who are good at exams' are not considered here, as the consequence of the effects being described by such arguments, namely that some people are selected for the wrong reason, is covered by the Validity criterion.

Compliance

9.48. Compliance in our terms is the extent to which the option prevents or mitigates the effects of cheating or malpractice. The scores for Security are presented in the table below.

Option	Compliance
Do Nothing	0
Multiple Mini Interviews	7
Structured Interviews	7
National Assessment for Ranking	10
SJT	10
Educational Performance	5

9.49. The rationale for the scoring is as follows.

9.50. Do Nothing is the only option in which the main inputs upon which selection is based are provided by the applicant in an uncertified way. Applicants could ask a third party to complete the answers, with little fear of detection. The current system attempts to detect such malpractice by sampling of forms and so on, but it remains, by a wide margin, the option most susceptible to malpractice.

9.51. Educational Performance ought to gain a high score, again on the grounds that it does not impose any additional test for selection purposes, so minimises the new opportunities for malpractice. That said, it may be difficult to police the scores that are submitted by all schools from outside of the UK, so for the time being the option has been marked down- the effects of this will be considered in the sensitivity analysis.

9.52. The interview options- being the only options in which the scorers are likely to know the applicant- is in theory open to deliberate bias, so they are marked down from the National Assessment and SJT.

Transparency

9.53. Transparency in these terms is the extent to which applicants can understand what is expected of them and why they have achieved one score rather than another. The scores for Transparency are presented in the table below.

Option	Transparency
Do Nothing	0
Multiple Mini Interviews	5
Structured Interviews	6
National Assessment for Ranking	7
SJT	3
Educational Performance	10

9.54. The rationale for the scoring is as follows:

9.55. Educational Performance does not involve any new tests, and is based on scores from existing assessments that should be well-understood by students. It is given the highest mark accordingly.

9.56. Feedback from applicants, gathered by MSC, suggests that many are unclear about the exact purpose of the white space questions, and about the qualities they should try to include in their answers in order to gain the best marks. In other words, applicants may be getting low marks because the exact nature of the test is not clear to them. For this reason it is given the lowest mark.

9.57. SJT is likely to be an unfamiliar concept to students, as it does not appear to be common in medical schools. Given this it is scored lower than the other remaining options.

9.58. The National Assessment is given a high score because the nature and purpose of knowledge-based assessments should be very familiar to students.

9.59. Structure interview is given a slightly higher mark than MMI because it is simpler, and therefore should be easier to understand.

Applicant Burden

9.60. Applicant Burden is the extent to which the option minimises the added cost and effort for applicants in addition what is required for their undergraduate course. The scores for this criterion are presented in the table below.

Option	Applicant Burden
Do Nothing	5
Multiple Mini Interviews	3
Structured Interview	3
National Assessment for Ranking	0
SJT	3
Educational Performance	10

9.61. The rationale for the scoring is as follows. Educational performance introduces no additional work for the applicant, since it is based on the assessments being made as part of the applicant's education in any event.

9.62. Do Nothing has two advantages for the applicant. Firstly they can complete their applications at a time and place of their convenience. More importantly, the application window is sufficiently long that they can see the questions and then put in the intellectual effort focused at those questions alone. With the other options the applicant does not know, in advance, what the questions would be, and so would have to prepare more widely. The National Assessment for Ranking is considered to be the worst option in this regard, by some margin.

Medical Time

9.63. This criterion measures the extent to which each option avoids the need for the time of medical staff from within then NHS. The scores for Medical Time are presented in the table below.

Option	Medical Time
Do Nothing	2
Multiple Mini Interviews	0
Structured Interviews	1
National Assessment for Ranking	0
SJT	10
Educational Performance	10

9.64. The rationale for the scoring is as follows.

9.65. Neither SJT nor Educational Performance impose any requirement for medical input, so they get the highest score.

9.66. The interview options and National Assessment have the highest requirement for medical time, and get the lowest score (whether one is worse than the other depends on the exact make-up of the interview panels and teams- this is considered in the sensitivity analysis).

9.67. Do Nothing requires significant input from clinicians for scoring.

Feasibility

9.68. Feasibility is a measure of the size and complexity of the management challenge that would be the consequence of selecting a particular option.

9.69. The difference in the feasibility of the options can be drawn out by breaking the management challenge down into the following elements:

9.70. **Rules-** how easy or difficult is it, from one year to the next, to gain agreement to the tests to be used, pilot them, write guidance, calibrate tests across schools and so on.

9.71. **Scorer logistics –** how easy is it to recruit scorers/interviewers, train them, organise their attendance at events.

9.72. **Applicant logistics-** how easy is it to provide guidance to applicants, get them registered with the process, book them into events, and so on.

9.73. **Technical-** how easy is it to maintain the national allocation system

9.74. One this basis the options can be judged as follows:

9.75. **Rules-** there is little to distinguish between the most of the options, except that a National Assessment for Ranking would meet with such professional hostility that there is a real possibility that consensus would never be reached. Note it could be argued that the Rules Effort associated with Educational Performance should be ignored, as it would simply be replacing work that would otherwise be done on developing curricula locally. It could also be argued that the Rules effort for Do Nothing in the short term is less arduous than the other options, as it is familiar ground for most of those involved.

9.76. **Scorer Logistics-** Educational Performance requires no additional effort. The other options each require large numbers of local testers to be recruited, trained, organised, etc. Do Nothing fairs marginally better in the short term as there will be people in the Foundation Schools familiar with the practice of scoring from previous years.

9.77. **Applicant Logistics-** Educational Performance and Do Nothing are best, as neither require the applicant to attend any specific event purely for the purpose of selection. National Assessment is the worst. The remaining options are on a par.

9.78. **Technical-** Educational Performance and National Assessment for Ranking are probably the easiest, in that, as envisaged in this document, they requires only a single score to be entered onto the system for each applicant. The remaining options are slightly more complex, depending on whether the interview booking is done on line, and whether a single score can be entered per applicant, or some breakdown is needed. .

9.79. In discussing the rationale for the scoring, it is important to consider the question of timing, as one option that seems most feasible in the long term might be entirely undeliverable within a short timescale. The two most significant factors here are that the Do Nothing option is an existing process, so is the most straightforward for the near future. If Educational Performance was to depend on changes within medical schools percolating through at least two years of an applicant's education, then it could not be implemented for at least the next three recruitment rounds. Given this long-term and short-term feasibility are scored separately as follows:

Option	Short Term Feasibility	Long term Feasibility
Do Nothing	10	3
Multiple Mini Interviews	4	3
Structured Interviews	5	3
National Assessment for Ranking	0	0
SJT	2	7
Educational Performance	0	10

Public Opinion

9.80. This is a measure of the extent to which the average 'person in the street' is likely to believe that the process is an acceptable one, without necessarily understanding its real merits. The scores for this criterion are presented in the table below.

Option	Public Opinion
Do Nothing	0
Multiple Mini Interviews	10
Structured Interviews	10
National Assessment for Ranking	10
SJT	5
Educational Performance	8

9.81. There is little direct evidence to support the scoring here. The rationale is as follows.

9.82. Interviews are an established traditional part of recruitment in many walks of life. The fact that they have been shown to be unreliable and invalid will not be known or appreciated by most people. Nor are people likely to realise that applicants are subject to all sorts of assessments and controls at medical schools before they are allowed to become doctors. Given this, it can expected that members of the public to assume that interviews would be part of the process.

9.83. The easily appreciated opportunity for the risk of cheating is likely to make Do Nothing the option that would gain least favourable public opinion, regardless of whether cheating actual happens or makes a difference to the selection outcomes.

9.84. National Assessment for Ranking conveys an impression of rigour and standardisation, so is likely to command respect, as would the idea of a score based on performance at medical school

9.85. SJT will not be well understood.

Weightings

9.86. The purpose of a weighting is to model the relative importance of the differences between the scores of the options in relation to one of the criterion. Note that this is quite different from weighting the criterion in absolute terms. For example, the strength of steel girders might be an extremely important criterion when building a bridge, but if all ones girders are sufficiently strong, then the minor differences between them become an unimportant criterion.

9.87. The weightings used for the base models, which were developed with the help of two members of the FPSG, are shown in the table below. Subsequently, the weightings were reviewed by other members of the FPSG, and by representatives of the Foundation Programme Rules Group- the significance of these reviews is considered in the sensitivity analysis.

9.88.

Name	Best Option	Worst Option	Weighting of the range
Reliability	SJT/MMI	Do Nothing	10
Validity	SJT	National Assessment	10
Granularity	SJT/National Assessment	Structured Interview	1
Consistency	SJT	Do Nothing	7
Longevity	Educational Performance	Do Nothing	12
Educational Impact	Educational Performance	National Assessment	15
Fairness	SJT	National Assessment	5
Compliance	SJT/National Assessment	Do Nothing	7
Transparency	Educational Performance	Do Nothing	2
Applicant Burden	Educational Performance.	National Assessment	4
Consultant Time	Educational Performance/SJT	MMI/National Assessment	8
Feasibility Short Term	Do Nothing	National Assessment	10
Feasibility Long Term	Educational Performance	National Assessment	10
Public Opinion	Structured Interview/MMI	Do Nothing	1

Weighted Performance

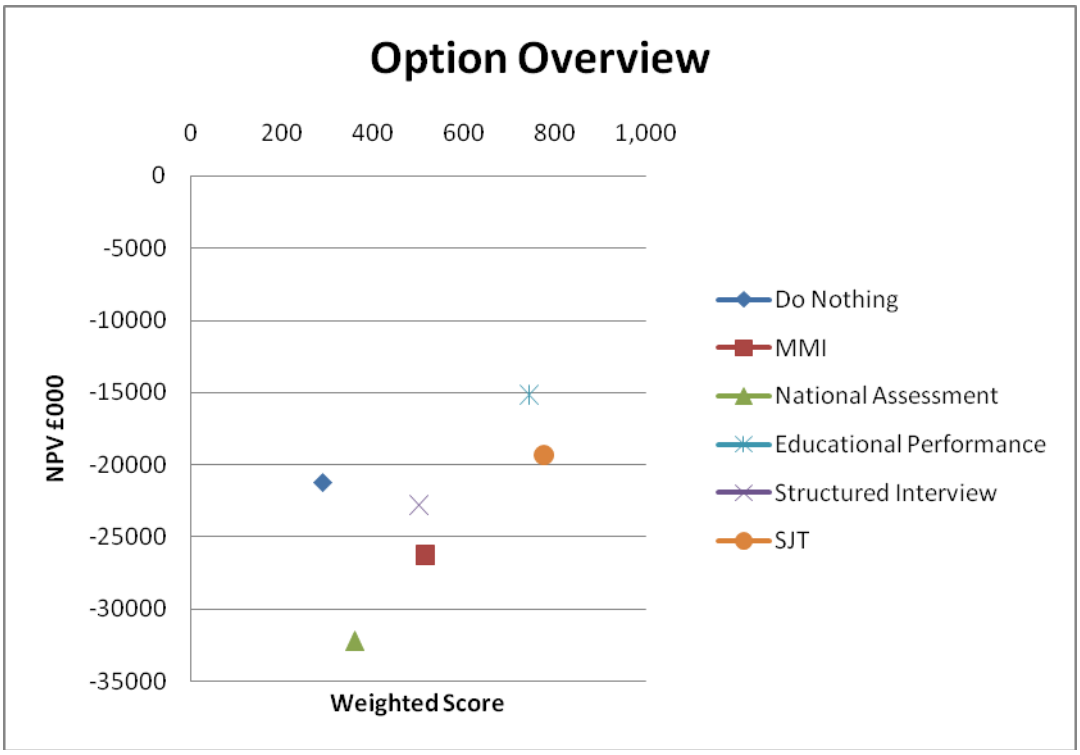
9.89. The result of applying the weightings to the scores is shown in the table below.

		Reliability	Validity	Granularity	Consistency	Longevity	Educational Impact	Fairness	Compliance	Transparency	Applicant Burden	Clinical Time	Short Term Feasibility	Long term Feasibility	Public Opinion
Do Nothing		0	3	5	0	0	4	7	0	0	5	2	10	3	0
MMI		10	8	5	2	8	4	4	7	5	3	0	4	3	10
Structured Interview		8	7	0	2	8	4	4	7	6	3	1	5	3	10
National Exam		10	0	10	8	9	0	0	10	7	0	0	0	0	10
SJT		10	10	10	10	9	6	10	10	3	3	10	2	7	5
Educational Performance		8	6	8	5	10	10	5	5	10	10	10	0	10	5
	Totals														
Raw weight	102	10	10	1	7	12	15	5	7	2	4	8	10	10	1
Scaled weight	100	10	10	1	7	12	15	5	7	2	4	8	10	10	1
Do Nothing	290	0	29	5	0	0	59	34	0	0	20	16	98	29	0
MMI	516	98	78	5	14	94	59	20	48	10	12	0	39	29	10
Structured Interview	501	78	69	0	14	94	59	20	48	12	12	8	49	29	10
National Assessment	361	98	0	10	55	106	0	0	69	14	0	0	0	0	10
SJT	775	98	98	10	69	106	88	49	69	6	12	78	20	69	5
Educational Performance	743	78	59	8	34	118	147	25	34	20	39	78	0	98	5

10. Comparisons and Sensitivity Analysis

Initial comparison

10.1. Based on the assumptions, scores and weightings applied so far in this document, the relative attractiveness of the options can be depicted as in the diagram below, in which the top right corner represents the most desirable outcome, and the bottom left the least desirable.



10.2. On this basis the most desirable options appear to be SJT and Educational Performance, as they have lower costs and higher weighted scores than all the other options.

10.3. Another way to view the results is to consider the ratio of the NPV to the weighted score, which can be thought of as a cost per benefit point. These ratios are shown the table below:

Option	Cost per Benefit Point
Do Nothing	£73,132
MMI	£50,916
National Assessment	£89,215
Educational Performance	£20,403
Structured Interview	£45,438
SJT	£24,888

10.4. However, this picture has arisen from making assumptions about the options, some of which may be uncertain. The purpose of the sensitivity analysis is to determine the extent to which plausible changes in the values of the assumptions could lead to a change in the rankings of the options.

Cost Insensitivities

10.5. The following observations arise straightforwardly from inspection of the cost estimates in Section 8.

10.6. Changes to the numbers of foundation schools and medical schools have a negligible impact on the costs.

10.7. The procurement costs are small compared with the overall costs. Possible variances in procurement costs by option (say plus or minus 50%) will not change the relative attractiveness of individual options.

10.8. Changes to the numbers of applicants scale the costs almost linearly; even if an extreme change is considered (say plus or minus 20%), the impact simply widens or narrows the spread between the cost of the options- it does not change the order of the options when ranked by cost.

10.9. Likely variances in the costs of piloting and developing the options (say plus or minus 50%)[10] will not change the order of the options when ranked by cost. Note that the costs are shown as arising in a single year; in practice the costs may be spread over a longer period, but again the effect of this does not change the overall outcome.

10.10. Likely variances in the costs of central administration are also unlikely to affect the ranking of the options by cost. Do Nothing and National Assessment are too widely separated in cost for this factor to make a difference. The two interview options are closer in terms of costs, but they are so similar that it is unlikely that the central administration costs will differ significantly. Educational Performance and SJT are less narrowly separated in cost terms; moreover, the difference between the two options is such that the central administration costs are likely to be less for the former than for the latter.

10.11. Likely variances in the 'rules' costs, by option, are too small to have a significant impact.

10.12. Travel expenses for applicants are a sizeable cost, but doubling or excluding them does not have a critical effect overall.

10.13. The possible variation in the cost of venues, between MMI and Structured Interviews, is likely to be small.

Dominant Cost Factors

10.14. The dominant uncertainties in cost models are as follows:

- The number of interviews that can be completed per interviewer day, and the associated opportunity cost per day for interviewers.
- The unit costs for National Assessment and SJT.
- The costs for implementing and maintaining the common content of assessment.

[10] The uncertainty here is quite large, particularly in relation to the differences between 'Do Nothing' and the other options. In 'Do Nothing' the re-procurement of the existing system is followed by a pilot of its successor. In the other options, the costs for piloting and development are all shown as being incurred before the contract for the existing service expires. In practice, some of the piloting and development costs will follow the re-procurement in these other options also.

10.15. These factors are considered in turn below.

Interviewer Costs

10.16. The largest single cost element for each of the two interviewing options is the cost of interviewer time. The assumptions that have been made in relation to these costs are not particularly conservative. The size of the panels for the structure interview (3 members) is not large, nor is the number of MMI stations (6). In practice these values could not be reduced to the point that either of the two interviewing options became a serious contender on cost. And since the options also score low in the non-monetary assessment, the cost uncertainty is not critical.

Unit Costs for Assessments

10.17. The GMC 'PLAB' costs have been assumed to be good analogues for the National Assessment and the SJT options. National Assessment is the more expensive of the two, because it is assumed to have an OSCE component (analogous to PLAB Part B), as well as a machine-markable knowledge test (analogous to the cheaper PLAB Part A), whereas SJT has only a machine-markable test. This is a critical assumption in cost terms- if the National Assessment could be made a machine-markable test, it would be broadly comparable in cost with SJT. Given that, however, National Assessment gains a low overall score for the non-monetary assessment, so SJT would appear to be a more reliably attractive option.

CCA costs

10.18. The costs for developing the common assessment items are included in all of the options, even 'Do Nothing', on the grounds that CCA appears to be going ahead for educational reasons regardless of whether it has a role to play in selection. If the CCA costs were completed excluded from the NPV calculations of those options for which it is not seen as important for selection- namely National Assessment and Do Nothing- then National Assessment would remain the most expensive option, but Do Nothing would become somewhat cheaper than SJT. However, given the poor performance of Do Nothing in relation to the non-monetary criteria, it is likely that SJT would still be seen as a more attractive option.

Scoring Uncertainties

10.19. Within Section 9 some provisional scores were awarded to some options on the grounds that the associated properties of the options were difficult to assess at this stage. These uncertainties are considered here, along with other issues related to scores and weighting that have been raised by stakeholders.

10.20. The Granularity of Educational Performance option is quite uncertain, but whether it is scored as 0 or 10 has a negligible effect on the overall position, so the uncertainty is not important.

10.21. The Consistency and Compliance scores of the Educational Performance option were difficult to predict, and were provisionally scored 5. If either score is switched to 0 or 10 the impact is a 5% change in the overall score for the option (assuming the weightings remain unchanged). These changes do not appear to be a decisive factor, given the overall uncertainty in the scores and weightings (see section on Conclusions, below).

10.22. The Fairness of Educational Performance was also difficult to predict. Switching its value to 0 or 10 has a 2% impact the overall score (assuming the weighting remains unchanged). This does not appear to be a decisive factor.

10.23. Structured Interview and MMI both were given a score of 0 in relation to the Medical Time criterion, as it was uncertain which would have the greater impact in practice. However, the uncertainties for these scores can be discounted, as these two options fall so far short of some of the others in terms of overall score.

10.24. In the review of the scores, one FPSG member queried the rationale for awarding National Assessment a low score for Feasibility, which was largely on the basis of hostility to the idea among key stakeholders. Changing the feasibility score can affect the overall score of the option by around 20% (given current weightings). However, the option is well adrift from the others in terms of its NPV, so the Feasibility score is not critical overall.

Weightings

10.25. The weightings used in the base model were compiled at a meeting with two FPSG members quite early in the work. Since then, other stakeholders have had an opportunity to comment on the weightings, and have independently suggested other weightings. Specifically, independent sets of weighting have been produced by:

- The DH representatives on the FPSG
- Representatives of the Foundation Rules Group (FRG), who determine the rules for the current selection method
- Attendees at the stakeholder events held on 16th June 2009.

10.26. There are considerable variations in the weightings that have been produced. These variations are doubtless in part due to different stakeholder perspectives. For example, the representatives of the FRG placed far less emphasis on 'Educational Impact' compared with weightings suggested by representatives of the Medical Schools Council and a medical student. However, the cumulative effect of these variations had little effect on the overall results. Specifically, regardless of whose weightings were used:

- 'Do Nothing' has the lowest score in all cases bar one (where it beats 'National Assessment')
- 'Structured Interview' Always scores less than 'MMI'
- 'SJT' and 'Educational Performance' always score much better than all of the other options, with SJT being marginally ahead in all cases.

10.27. This suggests that, while there is naturally scope to debate the individual scores and weightings, overall the 'SJT' and 'Educational Performance' options tend to perform better than the others, with SJT being marginally the better of the two.

10.28. The alternative weightings that have been suggested are shown in Annex A.

Optimism Bias and Risk

10.29. The dominant cost factors described above are also the areas that are most likely to be affected by optimism bias. The greatest scope for cost increases lie in those areas. In particular, it is quite possible that the interview productivity rates assumed in the base cost models could be too high, perhaps by as much as 30% say. That said, neither of those options would appear to be the most attractive, so the question is academic.

10.30. The more practical question is whether the costs assumed for Educational Performance or SJT are so optimistic that in reality neither of those two options would be cheaper than the current 'White Space' approach. Inspection of the estimates shows that the cost factors that are most uncertain would have to be out by a factor of 70-100% in order for that to be the case, which is unlikely.

10.31. The key risk that has not been taken into account in the modelling so far is that of a legal challenge. The cost of successfully defending a challenge (assumed to be the order of £150,000) is not significant when compared to the differences between the NPVs of the options. The higher impact risk is that of failing to defend a legal challenge. The options have yet to be subject to a legal review. On the face of the matter, however, the 'Do Nothing' option would appear to be riskier than Educational Performance or SJT, so the effect of that risk is to strengthen rather than reduce the case for change.

10.32. At this stage the most important risks to consider are those that might impact the options unequally, and thus might have a bearing on the relative attractiveness of the options. The two main risks that fall into this category relate to the availability of suitable interviewers- which applies to the two interviewing options and National Assessment- and to the professional antipathy expressed by the BMA towards the National Assessment option. The effect of these risks is to increase, rather than decrease, the relative attractiveness of Educational performance and SJT.

10.33. More generally, risks have been taken into account implicitly in the scoring of options against the assessment criteria. For example, the scores for 'Feasibility' take into account the likely difficulties associated with scheduling selection events.

Overall NPV Uncertainties

10.34. The sensitivity analysis above suggests that the uncertainties in the costs are unlikely to change the relative attractiveness of the two leading options compared with the rest. That said, for completeness it is interesting to consider what the overall levels of uncertainty within the NPV calculations might be.

10.35. The uncertainties in the costing assumptions fall into two groups, those that could affect all of the options, and those that can only affect some of the options. The key uncertainties in the first of these groups relate to the numbers of applicants, the costs of central admin, the average day rates for staff, and the common content of assessment. In the second group they relate to the duration of interviews, numbers of interviewers per panel, unit costs for the development of tests, and the costs of local administration. It is not trivial, or at this stage necessary, to model the cumulative impact of all these uncertainties in a detailed way. However, some broad observations may be made as follows:

10.36. The number of applicants has not varied significantly over the last three years, so the chance of a large variation in the near future would seem to be low. A 10% uncertainty seems sensible.

10.37. The main factor determining the day rates for staff is whether the medical specialists involved are consultants or less senior grades. A 25% uncertainty seems reasonable.

10.38. The central admin costs are affected in part by the number of applicants, and already contain a 40% uplift for optimism bias, which is itself a measure of uncertainty. Assume the costs could range from £1m per year to £1.5m.

10.39. The costs for the common content of assessment might vary by 30%, say.

10.40. The costs of interviews have been estimated on the basis of assumptions that reflect existing practice in medical recruitment, and are unlikely to be out by a large margin, but an uncertainty of 20-30%, say, might be reasonable.

10.41. The cost estimates for developing and delivering tests are based on actuals incurred by the GMC. The main uncertainties are in the extent to which the scope of the GMC test is comparable with what would be required for F1 selection, and whether deployment on a national scale would affect the unit costs. Again 20-30% uncertainty might be reasonable here.

10.42. The overall impact of the uncertainties described above depends upon their interaction. For example, a decrease in the unit costs of providing a test might be offset by an increase in the

cost of central administration. The chance of all of the uncertainties combining to have the same effect on the overall costs is low. Given this, and the observations above, it would be reasonable to assume an overall uncertainty in the NPVs of around 10% to 20%.

Conclusion

10.43. The options that have been considered in the report are defined in very general terms and at a high level, with considerable uncertainty about the exact way in which they might be implemented in practice. As a result, their NPVs are broad brush estimates, and there are uncertainties in the assumptions upon which the non-monetary analysis has been based. That said, the sensitivity analysis shows that this high level and uncertain picture is still sufficiently clear to allow conclusions to be drawn about the relative attractiveness of the options.

10.44. The National Assessment seems the least attractive of the options. On the plus side, it offers high levels of consistency, reliability and transparency, but so do some of the other options without the main drawbacks of the National Assessment, namely its high cost, its relatively low validity, its impact on educational objectives, its heavy use of clinical time, and the antipathy shown towards it by key stakeholders. There is also an issue of fairness with this option- is it right that the results of this one-off test of clinical skills and knowledge should be considered more important than an applicant's educational achievements over many years of study to become a doctor?

10.45. The Do Nothing option performs least well in non-monetary terms. Its main shortfalls are its relatively low validity and reliability, its high use of clinical time, and the fact that it is potentially open to mis-use by applicants.

10.46. The two interviewing options and SJT are similar to each other in that they combine some test of wider professional skills with an educational performance score that represents the applicant's clinical skills and knowledge as measured by their medical school. Of the three, SJT is distinctly more attractive- it out-performs the others in relation to almost all the assessment criteria, and is considerably less expensive.

10.47. SJT, then, appears to be clearly better than National Assessment, Do Nothing, and the two interviewing options. The remaining question is whether the educational performance score alone could be used for the purpose of selection, as envisaged in the Educational Performance (EP) option. Clearly such an approach would have some important advantages, which show up as a better NPV cost and high weighted scores for the EP option. However, the weighted scores for EP are based on two key assumptions. The first is that it will be possible for Medical Schools in the UK to agree a suitably robust, valid and fair approach to scoring the performance of their students, which could be reliably followed by non-UK medical schools; the FPSG is best-placed to judge that. The second assumption- which will be tested outside of the CBA work- that the burden imposed by the EP option on applicants and medical schools from outside the UK is acceptable legally, If either of these two assumptions is incorrect, then the total weighted score of the EP option would be substantially reduced, to the point at which its economic advantages would more clearly fail to offset its other shortcomings.

10.48. Given the above, the conclusions of the CBA work support the decision by the FPSG to take forward SJT and Educational Performance as possible options, and to seek legal opinions on the latter in particular. If the legality of the EP option were to be confirmed, then a short study to investigate the other key assumptions made about the feasibility of the option might be a sensible prelude to a pilot.

10.49. Finally, the choice of matching algorithm, which was described in Section 5, does not impact in any way the choice of selection option. One question is whether it is worth trading a concrete practical advantage of the current hybrid algorithm- namely a marginally higher proportion of applicants getting their declared first choice- for the greater fairness of the pure 'serial dictator' algorithm. Subject to legal advice on the issue of fairness, it would seem that the applicants

themselves, as the stakeholder group with the greatest interest in the question, should have a say in the answer.

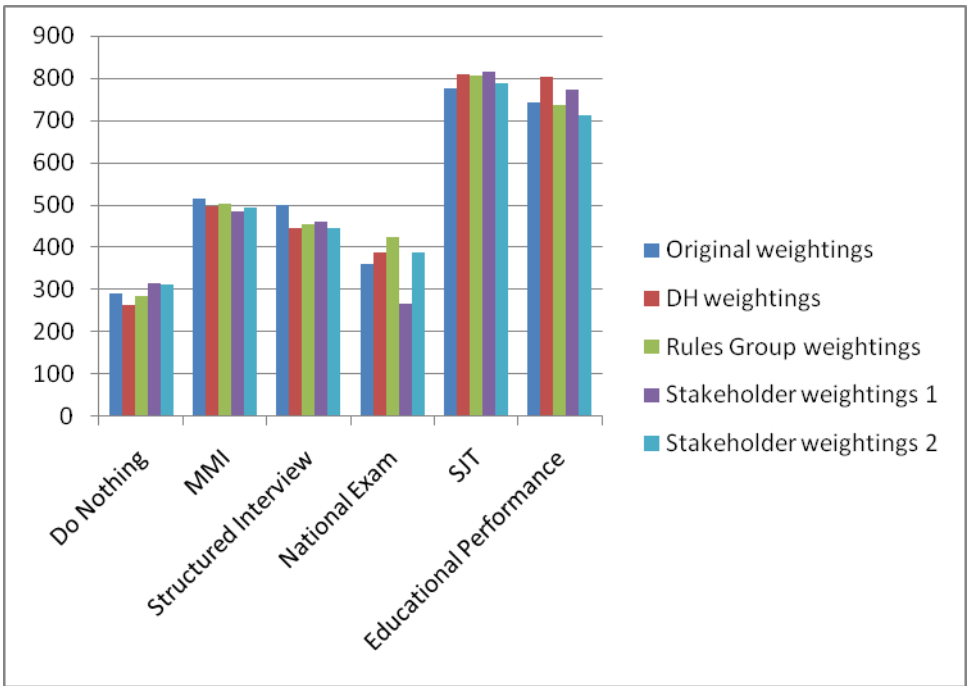
11. Document History

Version	Date	Owner	Status	Changes
0.1	8/6/09	D.Shaughnessy	Initial Draft	
0.2	9/6/09	D.Shaughnessy	Draft for FPSG review	Update following QA review with MSC. Situational Judgement Test relegated to sub-option. Criteria, scores and weightings updated. Some changes to terminology.
0.3	14/6/09	D.Shaughnessy	Draft for FPSG review	General update following QA comments from FPSG members. Significant extension of the section on sensitivity analysis. SJT restored to full option status. Structured Interview added as an option.
0.4	30/6/09	D.Shaughnessy	Issued for FPSG consideration	General update following QA comments from FPSG members, and discussion at FPSG meeting on 26/06/09. Scorings and weightings updated. Costs for common content of assessment added to all options bar National Assessment.
0.5	2/7/09	S.Fitzpatrick	Issued for FPSG consideration	Changes following QA reviews by S.Fitzpatrick and P.O'Neill. 'White Space' option updated and re-labelled as 'Do Nothing'; definition of 'Educational Performance' option clarified; additional commentary in the sensitivity analysis section; corrections to some NPV figures; a change to the scores for 'Consistency'; introduction of Annex A.
0.6	4/8/09	S.Fitzpatrick	Issued	Changes following QA discussion with J. Perkin on 15/7/09. Cost of 'do-nothing' pilot reduced; an optimism bias cost has been built in; the rules costs for SJT have been increased slightly; some clarifications have been added in Sections 8 and 10.

12. Annex A- Alternative Weightings

This annex shows the effects of alternative weightings that have been suggested by DH, the Foundation Rules Group, and participants at the stakeholder workshop on 16th June 2009.

The effects are summarised in the graph below, which shows how the total weighted scores of each options varies according to which set of weightings is used. It is clear from the graph that the SJT and Educational Performance options score far more highly than the rest regardless of which set of weightings is used.



Details of the weightings are provided on the following pages.

DH Weightings

		Reliability	Validity	Granularity	Consistency	Longevity	Educational Impact	Fairness	Compliance	Transparency	Applicant Burden	Clinical Time	Short Term Feasibility	Long term Feasibility	Public Opinion
Raw weight	132	15	15	9	9	9	15	5	9	3	15	15	3	9	1
Scaled weight	100	11	11	7	7	7	11	4	7	2	11	11	2	7	1
Do Nothing	263	0	34	34	0	0	45	27	0	0	57	23	23	20	0
MMI	498	114	91	34	14	55	45	15	48	11	34	0	9	20	8
Structured Interview	445	91	80	0	14	55	45	15	48	14	34	11	11	20	8
National Exam	389	114	0	68	55	61	0	0	68	16	0	0	0	0	8
SJT	810	114	114	68	68	61	68	38	68	7	34	114	5	48	4
Educational Performance	805	91	68	55	34	68	114	19	34	23	114	114	0	68	4

FRG Weightings

		Reliability	Validity	Granularity	Consistency	Longevity	Educational Impact	Fairness	Compliance	Transparency	Applicant Burden	Clinical Time	Short Term Feasibility	Long term Feasibility	Public Opinion
Raw weight	93	10	10	7	8	5	3	10	6	8	6	7	5	8	0
Scaled weight	100	11	11	8	9	5	3	11	6	9	6	8	5	9	0
Do Nothing	285	0	32	38	0	0	13	75	0	0	32	15	54	26	0
MMI	502	108	86	38	17	43	13	43	45	43	19	0	22	26	0
Structured Interview	454	86	75	0	17	43	13	43	45	52	19	8	27	26	0
National Exam	425	108	0	75	69	48	0	0	65	60	0	0	0	0	0
SJT	808	108	108	75	86	48	19	108	65	26	19	75	11	60	0
Educational Performance	738	86	65	60	43	54	32	54	32	86	65	75	0	86	0

Stakeholder Meeting Weightings 1

		Reliability	Validity	Granularity	Consistency	Longevity	Educational Impact	Fairness	Compliance	Transparency	Applicant Burden	Clinical Time	Short Term Feasibility	Long term Feasibility	Public Opinion
Raw weight	189	15	25	3	9	9	25	25	15	3	15	15	5	25	0
Scaled weight	100	8	13	2	5	5	13	13	8	2	8	8	3	13	0
Do Nothing	315	0	40	8	0	0	53	93	0	0	40	16	26	40	0
MMI	484	79	106	8	10	38	53	53	56	8	24	0	11	40	0
Structured Interview	459	63	93	0	10	38	53	53	56	10	24	8	13	40	0
National Exam	267	79	0	16	38	43	0	0	79	11	0	0	0	0	0
SJT	815	79	132	16	48	43	79	132	79	5	24	79	5	93	0
Educational Performance	772	63	79	13	24	48	132	66	40	16	79	79	0	132	0

Stakeholder Meeting Weightings 2

		Reliability	Validity	Granularity	Consistency	Longevity	Educational Impact	Fairness	Compliance	Transparency	Applicant Burden	Clinical Time	Short Term Feasibility	Long term Feasibility	Public Opinion
Raw weight	44	5	5	3	5	1	3	5	1	3	3	3	3	3	1
Scaled weight	100	11	11	7	11	2	7	11	2	7	7	7	7	7	2
Do Nothing	311	0	34	34	0	0	27	80	0	0	34	14	68	20	0
MMI	493	114	91	34	23	18	27	45	16	34	20	0	27	20	23
Structured Interview	445	91	80	0	23	18	27	45	16	41	20	7	34	20	23
National Exam	386	114	0	68	91	20	0	0	23	48	0	0	0	0	23
SJT	789	114	114	68	114	20	41	114	23	20	20	68	14	48	11
Educational Performance	714	91	68	55	57	23	68	57	11	68	68	68	0	68	11