**Research Report
No 290**

department for
education and skills

Feasibility Study for the Long Term Evaluation of Modern Apprenticeships

Joan Payne – Policy Studies Institute
Rebecca Riley – National Institute of Economic and Social Research
Nick Coleman – BMRB International

Acknowledgements

Disclaimer

CONTENTS

6 A NEW SURVEY OF MA ENTRANTS

7 OTHER RESEARCH AIMS

8 TIMEFRAMES AND OPTIONS

ABBREVIATIONS

AMA	Advanced Modern Apprenticeship
DfES	Department for Education and Skills
DfEE	Department for Education and Employment
EMA	Education Maintenance Allowance
FMA	Foundation Modern Apprenticeship
GCSE	General Certificate of Secondary Education
GNVQ	General National Vocational Qualification
IPR	Individualised Pupil Record
JUVOS	Joint Unemployment and Vacancies Operating System
LEA	Local Education Authority
LFS	Labour Force Survey
LSC	Learning and Skills Council
MA	Modern Apprenticeship
NDYP	New Deal for Young People
NEET	Not in education, employment or training
NES	New Earnings Survey
NINO	National Insurance Number
NTr	National Traineeship
OT	Other Training
SAT	Standard Attainment Test
TDS	Trainee Database System
TEC	Training and Enterprise Council
VET	Vocational education and training
WBTYP	Work-based Training for Young People
YCS	Youth Cohort Study
YTS	Youth Training Scheme
YT	Youth Training

SUMMARY

Introduction

This report investigates the feasibility of conducting a formal evaluation of Modern Apprenticeships (MA) with two main aims:
(a) to produce robust quantified estimates of the additional (net) impact of MA on trainees' long-term job chances and earnings;
(b) to examine the benefits of MA for employers or for the wider economy, and variations in the effectiveness of MA between different sectors of the economy.
The DfES required recommendations on the best methodology to produce results by four different dates, September 2001, January 2002, September 2003 and September 2005. Should it be impossible to meet aims (a) and (b) by these dates, recommendations were required for less rigorous studies that nevertheless add to knowledge about outcomes from MA.

The feasibility study was carried out in the space of a very few weeks. It has benefited from discussions with an advisory group, by a workshop attended by academics, research professionals and policy experts with an interest in the area, and from written comments on an interim report.

Development of Modern Apprenticeships

The design of the proposed evaluation study will be influenced by the way the MA programme has developed since it was launched nationally in 1995. It is often inadvisable to base a formal evaluation of a programme on the earliest entry cohort, which may be affected by teething troubles. To include Foundation Modern Apprenticeships (FMA) as well as Advanced Modern Apprenticeships (AMA), the first entry cohort that the evaluation could be based on is 1998-99.

Both AMA and FMA have drawn young people away from Other Training (OT) within Work-Based Training for Young People (WBTYP), and the characteristics of MA entrants are likely to have changed year by year as OT has declined.

The progressive expansion of MA into new sectors means that an evaluation based on an early entry cohort would be biased towards the more traditional sectors in which MA was first launched. As the demographic profile of MA trainees varies between sectors, the expansion into new sectors has also changed the overall demographic composition of MA trainees.

Reforms to MA currently under discussion have two strands: first, a drive to raise standards, increase retention and provide more opportunities for progression, and second, to widen access and draw in more socially excluded young people. These developments are likely to lead to further changes in the characteristics of MA entrants in future years.

Strategies for estimating net impacts on individuals

To measure the net impact of a social programme on participants, we must estimate the counterfactual, namely what would have happened to them if they had not taken part in the programme. The main problem is that participants usually differ systematically from non-participants in ways that affect the likelihood of a successful outcome. With MA, differences are found both when comparing MA participants as a whole with young people taking other routes, and when comparing trainees on different strands within the MA programme. In evaluation research several different strategies are used to deal with such selection effects.

Random allocation designs offer perhaps the soundest methodological approach. However ethical, practical and political obstacles mean that there is no possibility of using such a design to evaluate MA.

The 'instrumental variable' approach relies on finding variation in the probability of participating in the programme that is unrelated to the probability of a successful outcome other than through its effect on participation. This variation can sometimes be artificially created. This strategy could perhaps be applied to MA through randomly allocated financial incentives to train, though it is likely to face political opposition. It is hard to identify a naturally occurring instrument for participation in MA.

Matched area designs are often used in the UK to assess the impact of a new programme that is launched in a limited number of pilot areas. There is no possibility of using this strategy to evaluate MA, as the programme has been running nationally for several years and cannot be selectively withdrawn.

The 'difference in differences' strategy proved valuable in evaluating the impact of the New Deal for Young People (NDYP). However, unlike NDYP, MA does not have a well-defined target group and its launch did not mark a significant change from previous policies, so this approach is unlikely to yield clear results.

Matched comparison group designs compare the progress of programme participants with that of eligible non-participants who are similar to participants on observable

factors that affect both the chances of participation and the chances of a successful outcome. The aim is to reduce any unobserved differences between the participant and non-participant samples to a minimum by matching on a very wide range of variables. The appropriateness of this approach depends on the strength of selection into the programme, but in the case of MA it appears in principal to be a viable strategy. The main difficulty lies in finding a suitable data set from which to construct the matched groups, or a suitable sampling frame for a new survey of a comparison group.

Simple statistical modelling without prior matching but with a wide range of control variables is sometimes used where a data set is already available for analysis, but may lead to biased estimates of programme effects.

Specific issues in evaluating MA

The choice of an appropriate comparison group for MA trainees is by no means obvious. Different comparison groups largely correspond to different parameters of interest, which answer different policy questions. There is potential policy interest in comparisons with full-time vocational education, with WBTYP outside the MA framework, with jobs outside of WBTYP, especially in sectors where MA has only recently been established, and with young people not in education, employment or training. Ideally the evaluation study would include more than one comparison group.

Finding a suitable sampling frame for a new survey of a comparison group presents problems.
- For those in OT, the Trainee Database System (TDS) could be used.
- Sampling young people in full-time vocational education currently relies on the co-operation of schools and colleges and may not be very successful. It should become easier when the Individualised Pupil Record (IPR) is introduced.
- Young people in jobs outside of WBTYP would have to be identified through their employers, and the level of co-operation may be low. In some occupations, nearly all young people enter through WBTYP.
- There is no sampling frame for young people not in education, employment or training (NEET). The NEET group is still problematic if an existing data set is used to identify comparison groups.

Determining the age at which outcomes are measured will partly depend on the choice of comparison group. We must also take into account the fact that entrants to MA span the full 16-24 age range, that AMA entrants are on average older than entrants to FMA, and that people starting MA training in the traditional craft sectors tend to be younger than those starting training in the newer sectors.

A prospective evaluation design is stronger than a retrospective design, but takes longer to produce results. A retrospective design could be based on either a cohort of entrants to MA or a cohort of leavers, the former being more useful. Because of the wide spread of ages on entry to MA, basing the evaluation on a new age cohort of young people means a very long wait for results. Existing age cohort studies of young people give only partial coverage of MA because the last follow-up tends to take place at too young an age.

The evaluation study should measure drop-out from MA and estimate the impact of MA on young people who do not complete their training. Separate estimates should be produced for AMA and FMA. Data on the local economy should be included.

Possible data sources

Data already available for secondary analysis

The Youth Cohort Study (YCS) has the attractions of a reasonable set of variables to control for selection effects, good questions on post-16 education and training, and the possibility of defining more than one comparison group. Cohort 9 is the earliest cohort that could be used, though sample numbers for AMA entrants are on the margins of viability after allowing for sample attrition. The analysis would be restricted to 16 year old entrants, making it difficult to generalise from the results, and could not cover the impact of AMA on earnings. FMA entrants could not be included in the evaluation at all.

The Labour Force Survey (LFS) could only be used cross-sectionally, to compare people who had completed MA with others. Sample numbers are adequate for this, and there is very good information on earnings and employment. However there are several problems in using the LFS on its own.
- People who leave MA before completion are not identified.
- AMA and FMA are not distinguished from each other.
- There is no information on training sector, dates of entering and completing MA or geographical location at entry.
- Only a limited set of variables is available to control for selection effects.
- The choice of comparison group is restricted by the lack of information on education, training and work histories.

However it could be possible to use the LFS to provide comparative data for estimates of the employment rates and earnings of MA trainees derived from other sources.

The Trainee Database System (TDS) covers all WBTYP participants and has good data on the planned training programme at the point of entry. Variables available to

take account of selection effects are fairly limited. Information on outcomes largely relies on a postal survey six months after leaving which has a very low response rate. There is however a theoretical possibility of linking TDS data with JUVOS data on claims for unemployment benefit and with data on earnings from the New Earnings Survey (NES). Some comparisons could be made with groups identified in the LFS.

Surveys now in progress

YCS Cohort 10 suffers the same general drawbacks as Cohort 9. It has more respondents in FMA but fewer in AMA, and as yet it is not known whether a sweep at age 19/20 will be carried out. The earliest that this could take place is autumn 2002.

Cohort 1 of the Educational Maintenance Allowance (EMA) pilots study is the same age as YCS Cohort 10 and a sweep is planned in the fourth year after the end of compulsory education. As with YCS, using the EMA data for the evaluation would mean excluding older entrants to MA. Sample numbers for AMA and FMA are of a similar general order to YCS 10, but the data are likely to be richer and of better quality. However the EMA sample is not nationally representative, and this creates difficulties in generalising from results on MA.

It would be possible to conduct a limited evaluation comparing data from YCS Cohort 10 and the EMA pilots.

Possible future data sources

The DfES's proposed new longitudinal survey of young people potentially offers the richest and best quality data for an evaluation of MA. Of all the data sources examined, it is the best suited to applying the matched comparison group methodology. Sample numbers for MA trainees would be adequate. However the proposed survey will not yield suitable data until 2006 at the earliest, and including older entrants to MA in the evaluation would entail an even longer wait for results.

It might be possible to improve LFS data on MA by modifying certain of the survey questions. However competition for space in the LFS questionnaire casts doubt on the viability of this option, and the value of the new data would be limited by other problems.

A later sweep of YCS Cohort 9 or 10 when respondents are in their early twenties would allow older entrants to MA to be included in the evaluation, but high sample attrition would probably make this unviable.

A new survey of a cohort of WBTYP entrants drawn from the TDS could significantly enhance the data already available on outcomes from MA. Although it would not fully meet the DfES's aims for the evaluation study, it seems to be the best option that is practical given the timescales within which results are required. It would allow the net impact of MA to be assessed relative to OT, and outcomes for those who complete their training to be compared with outcomes for those who leave early. In addition, survey questions could be modelled on the LFS, so that direct comparisons could be made with the employment rates and earnings of groups identified in the LFS who had not completed apprenticeships. The viability of a new survey is explored in the next section.

A new survey of MA entrants

The proposed new survey would be based on a cohort of entrants to AMA, FMA and OT trainees drawn from the TDS. The cohort would be drawn on the basis of entry rather than leaving date to permit consistent analysis of outcomes, and would include entrants of all ages and from all sectors. An entry cohort covering a full year would be best, to balance out any seasonal variations in entry patterns. We envisage a cohort entering either between September 1998 and August 1999 or between September 1999 and August 2000. An achieved sample size in the region of 4,000-5,000 is likely to be required if separate analyses are to be conducted for AMA, FMA and OT. However, if separate analyses are not required, the sample size and associated costs can be reduced substantially.

The sample should be interviewed at least three years after entry. This length of time would be required to allow participants to complete their training and for sufficient time to elapse for meaningful analysis of outcomes. A face-to-face survey would be necessary, in order to cover the detail required, and to achieve sufficient response rates. Analysis and reporting would take a further 6 to 12 months.

A key issue for the survey would be sample attrition. This would need to be contained to a reasonable level, so that the achieved sample was sufficiently robust and representative. We outline a number of measures to tackle attrition, including pre-survey exercises to assess the likely level of attrition, methods of updating database information, and techniques while the survey is in progress. If sample attrition could be contained, a reasonable response rate could be achieved, given the relevance of the survey to respondents.

Other research aims

In addition to the net impact of MA on participants' long-term job chances and earnings, the DfES wish to consider the feasibility of evaluating the long-term benefits of MA from the perspective of employers or the wider economy and of measuring variations in the effectiveness of MA between different sectors of the economy.

In principle, an evaluation of the benefits of MA to employers, in terms of improvements in productivity and profitability, could be conducted in much the same manner as an evaluation of the benefits of MA to individuals, in terms of improvements in earnings. However, given the difficulties involved in obtaining data on firm characteristics, productivity and profitability, the feasibility of conducting an evaluation of this kind is limited and is not recommended. These data difficulties are further compounded by the evaluation problem itself. A comparison group of employers is difficult to obtain given that for a number of occupations MA is the predominant mode of entry and given the relatively high sectoral concentration of MA.

To the extent that MAs make a net difference to the productivity and employability of participants, the policy will have wider economic implications. An evaluation of the benefits of the policy as a whole requires an assessment of these, and will provide information on the benefits of the policy to employers. An assessment of the wider implications is also useful in informing the evaluation of the policy's impact on participants' earnings and long-term job chances, as the wider implications of the policy in turn affect both participants and the comparison group.

It is clear that the introduction of MA is likely to affect productivity and growth, earnings, employment and competitiveness, if it is successful in raising participants' productivity and participation. These effects are mostly long-term.

Generally, one can attempt to quantify the effects of policy either by econometric investigation or by calibrated model simulation of an internally consistent and theoretically coherent model. For the purposes of evaluating the wider economy effects of MA, relying on econometric investigation alone is unlikely to be a fruitful evaluation strategy. In terms of the aggregate economy, the effects of MA particularly in the early years will be too small to detect in this manner. Instead an evaluation will have to be conducted *ex ante*, in the form of an appraisal, relying on model simulation informed by empirical investigation.

The outcome of an *ex ante* evaluation of the wider economic effects of MA will be very sensitive to the way in which the evaluator perceives the relation between the

policy and the macro economy. For the purposes of evaluating MA, the crucial relationships are the factors influencing the participation decision and the skill-specific wage determination process, the extent to which there may be spillovers from MA training and the sectoral distribution of productivity changes in terms of openness to trade. A number of alternatives should be presented in order that the range of potential outcomes can be discussed.

An important input into an *ex ante* evaluation of the impact of MA on the wider economy is an estimate of the individual productivity gain associated with MA participation. This is best obtained from an evaluation of the impact of participation on individuals' earnings, but may also be obtained from the literature on the returns to training.

Differences in the effect of MA across relatively broad sectors such as manufacturing and services could be facilitated in the evaluation of the impact of MA on individuals. It would not be feasible to estimate sectoral differences at a more disaggregate level, due to the occupational and industrial concentration of MA trainees. These sectoral distinctions are also useful in evaluating the impact of MA on the wider economy.

Timeframes and options

The following options are available within the timeframes specified by the DfES.

Measuring the net impact of MA on job chances and earnings:
Results by September 2001: Preliminary analysis of the LFS or the TDS.
Results by January 2002: More thorough analysis of the LFS or TDS using more sophisticated statistical techniques, plus analysis of YCS Cohort 9 Sweep 4.
Results by September 2003: Analysis of YCS Cohort 10 and the EMA Pilots data for Cohort 1; analysis of the LFS with expanded questions on apprenticeship; analysis of a 21/22 year old sweep of YCS Cohort 9; new survey of a cohort of WBTYP entrants.
Results by September 2005: New survey of a cohort of WBTYP entrants.
Results by September 2007: Analysis of the new longitudinal study of young people.

Estimating the net impact of MA on the wider economy
Results by September 2001: Literature survey and simple analysis within pre-existing model of the UK economy
Results by January 2002: Analysis within basic long-run equilibrium model of the UK economy designed for an evaluation of the wider effects of MA
Results by September 2003: Analysis within fully specified long-run equilibrium model of the UK economy designed for an evaluation of the wider effects of MA

Results by September 2005 and 2007: As for September 2003, but using increasingly reliable information on the direct effect of MA participation on individual productivity.

1 INTRODUCTION

1.1 Aims

The Department for Education and Skills (DfES) commissioned a consortium consisting of the Policy Studies Institute, the National Institute for Economic and Social Research and BMRB International to investigate the feasibility of conducting a formal evaluation of Modern Apprenticeships (MA). The aims of this proposed evaluation were:

(a) to produce robust quantified estimates of the additional (net) impact of MA on trainees' long-term job chances and earnings;

(b) to examine other related issues, including the benefits of MA for employers or for the wider economy, and variations in the effectiveness of MA between different sectors of the economy.

The DfES asked for recommendations about the best methodology for achieving the evaluation aims, given constraints of timing and available data. More specifically, recommendations were requested on the best options for producing results within four timeframes of varying length:
- by September 2001,
- by January 2002,
- by September 2003,
- by September 2005, and forward in two-yearly cycles.

These dates were determined by the anticipated information needs of the government-wide spending reviews.

1.2 Constraints

The DfES accepted that it might be impossible to meet the aims of the proposed evaluation within the shorter of these timeframes, and in this case it required the feasibility study to suggest alternative studies that could be carried out more quickly. While they might not meet the rigorous standards of evidence needed to measure the additional impact of MA on long-term job chances and earnings, the DfES specified that any alternative studies should nevertheless make a useful contribution to knowledge about outcomes from MA.

The potential impact of MA on individuals is not, of course, limited to its impact on job chances and earnings - it could also have benefits whose economic value is not directly measurable, such as increased job satisfaction. However, in view of the short time available for the feasibility study, it was required to confine itself to the topics stated above. It was also beyond the scope of the study to explore the possibility of conducting a full cost-benefit analysis of MA, though of course meeting aims (a) and (b) is an essential first step towards such an analysis.

1.3 Method of working

The DfES's timetable allowed only a very few weeks for the feasibility study. The first step was to set up an advisory group consisting of research and policy specialists from the DfES and two external experts in evaluation methodology. Following discussions in the advisory group and further investigations, an interim report was drawn up which was circulated both within and outside the DfES. The interim report was the focus of a subsequent workshop attended by a number of academics, research professionals and policy experts with an interest in the area. The final report has been modified in the light of both discussion at the workshop and written comments on the interim report.

1.4 Structure of the report

The report begins with a review of the way that MA has developed since its launch, before turning to a discussion of methodological issues in its evaluation. Although the twin aims (a) and (b) of the proposed evaluation are inter-linked, it is convenient to discuss them separately. Thus how to measure the additional impact of MA on participants' long-term job chances and earnings is discussed in Sections 3-6, and the other issues raised under aim (b) are discussed in Section 7. Section 8 summarises the options available within the specified timeframes.

In writing the report, we have assumed that the intended audience is already familiar with the MA system, and so we have not felt it necessary to explain the details of its operation or to define commonly used terms.

2 DEVELOPMENT OF MODERN APPRENTICESHIPS

Since it was launched nationally in 1995, the MA programme has undergone considerable development, and consultations have recently been completed on further reforms.[1] These developments have implications for design of the evaluation study, and so this section is devoted to a brief review of these.

2.1 AMA, FMA and OT

In its first three years, the MA programme only covered what are now known as Advanced Modern Apprenticeships (AMA), originally termed simply Modern Apprenticeships, which are designed to lead to a Level 3 qualification. The number of starts on AMAs rose rapidly, reaching 83,300 during the year 1997-98 and remaining around this level during the following two years.[2] Any new programme tends to have teething troubles in its early stages, and it is generally advisable to defer formal evaluation until these have been sorted out and the programme has reached a stable form.

In the case of MA, there is an additional reason for avoiding basing the evaluation on those who did their training in the programme's earliest years. The initial rapid growth in MA starts was achieved partly by allowing young people in other WBTYP to transfer to AMAs, and it has been suggested that young people transferring in this way were not always properly qualified to do an AMA or suited to doing one. If so, the progress of early AMA trainees may be misleading, and indeed the proportion of young people completing their AMA training who gained a full qualification at Level 3 or above rose from an estimated 30% in 1995-96 to an estimated 54% in 1997-98.[3]

Foundation Modern Apprenticeships (FMAs), originally termed National Traineeships (NTr) and designed to lead to a Level 2 qualification, were launched two years after AMAs, with the number of starts again rising rapidly to reach 86,600 during 1999-2000.[4] As with AMAs, initially a number of young people transferred to FMAs from other WBTYP.

Since the start of the MA programme, the number of young people in WBTYP outside the MA framework, now referred to in official statistics as 'Other Training'

---

[1] See DfEE 2001.
[2] DfEE 2000b, Table 1.
[3] Middlemass 1999, Table 1 (estimates based on the TDS postal follow-up survey).
[4] DfEE 2000b, Table 1.

(OT), has declined very quickly. In 1999-2000 there were fewer starts in OT than in either AMAs or FMAs, and by September 2002 recruitment to OT will be ended and other arrangements put in place for young people who are not deemed ready for an apprenticeship. As both the better training programmes and the most able and motivated young people tend to be absorbed soonest within the MA framework, the gap in standards between the MA programme and OT can be expected to have widened as OT has declined.

2.2 Expansion into new sectors

MAs were first introduced in those industrial sectors that have a long tradition of apprenticeship training such as engineering and hairdressing. Since then, MA frameworks have been developed in a much wider range of industries, including service industries and IT-based industries. For AMA there are now 81 approved sectors, with 50 sectors approved for FMA. This again suggests it would be inadvisable to base an evaluation on an early entry cohort, as this would not provide any information about outcomes from training in the newer sectors.

The demographic profile of MA trainees varies considerably between sectors. For example, trainees in engineering manufacturing are nearly all white males and hairdressing takes in mostly white females. Childcare, while largely female, has an above average proportion of ethnic minority trainees. There are also variations between sectors in the age of entrants. Trainees in engineering, construction and hairdressing tend to enter at age 16, while trainees in retailing and other service sectors tend to enter at a slightly older age, while legal requirements mean that apprenticeships in certain sectors (for example butchery) cannot begin until at least age 18. Thus the development of MA frameworks in new sectors has been accompanied by changes in the overall demographic composition of MA trainees. The proportion of women entrants has increased, as have the proportions of Black, African or Caribbean entrants and of Asian entrants, while the proportion of entrants with disabilities has fallen.

2.3 Future developments

Proposed future reforms to MA have two main strands.

First, there is a drive to raise standards and thereby to improve levels of retention and to provide more opportunities for trainees to progress to higher levels of training and qualification. It is planned to achieve this through the introduction of technical certificates, an apprenticeship diploma, increased provision for the development of

Key Skills, a new inspection system for training providers, guidance on number of hours spent in off-the-job training and minimum recommended periods of training.

The second strand of the proposed reforms is about widening access, and includes a commitment to provide MA opportunities to all who can benefit from them, with an entitlement to MA training for all those with the necessary aptitude and ability. There is a particular desire to draw into MAs young people who are currently not in education, employment or training – the so-called 'NEET' group. It is hoped to achieve this by improving and extending the Learning Gateway, by providing pre-apprenticeship training, and possibly by the use of targeted financial incentives.

Currently MA trainees tend, as a group, to have better Year 11 GCSE results than either young people in jobs without training or the NEET group, but poorer results than those who stay on in full-time education after age 16. In addition, within the MA framework, there are differences between trainees in AMAs and FMAs, with the former having on average better GCSE results. The plans to raise standards and widen access are likely to change this picture, perhaps reducing the difference in average levels of school attainment between MA trainees and other young people not in full-time education, while increasing the diversity of levels of school attainment amongst MA entrants. There may be similar changes in motivational and attitudinal factors.

3   STRATEGIES FOR ESTIMATING NET IMPACTS ON INDIVIDUALS

3.1 The problem of selection effects

In order to measure the net impact of any programme, we must estimate the counterfactual case, namely what outcomes for participants would have been if they had not taken part in the programme. By definition it is impossible to observe this directly, so some kind of proxy measure must be constructed. In the evaluation of social programmes, the main problem to be faced in estimating the counterfactual is that of selection effects. If the programme is voluntary, participants will be self-selected, and if places are limited, there will be selection by those who allocate places. In addition there may be unintentional selection mechanisms that operate indirectly via the structure of the programme or the way that it operates. Selection effects can involve factors that are easy to observe and to measure, such as age and qualifications, and also factors which are difficult to observe and measure, such as attitudes and personality. It follows that in evaluating how successful a programme is in achieving its intended outcomes, it is not enough simply to compare what happens to participants with what happens to people who were eligible for the programme but did not take part, as in doing this we are unlikely to be comparing like with like.

In the case of MA, selection effects have been observed on sex (a majority of starts on AMAs are male, a majority of starts on FMAs are female), ethnicity (members of ethnic minorities are under-represented on AMAs), region (Greater London has a comparatively low number of starts) and prior qualifications (trainees in AMAs have on average poorer GCSE results than young people who stay on in full-time education, but better results than those in OT, in jobs outside the MA framework or with no job or training place).[5] Compared to young people who stay in full-time education to study A levels or GNVQs, MA trainees tend to have less favourable attitudes towards school and more experience of part-time and casual work before age 16.[6] Their attitudes are also likely to differ from those of young people in jobs outside the MA framework, as they must study for qualifications and often accept a lower wage initially than they could earn elsewhere.[7]

The basic problem with selection effects is that many factors that predict programme participation also predict programme outcomes. In the present case, many of the factors that are associated with doing an MA are also associated with future

---

[5] DfEE 2000b; Payne 1998; Payne 2001.
[6] Coleman and Williams, 1998.
[7] Payne 2001, Table 2.8.

employment chances and earnings, either negatively or positively, depending on the group with whom MA trainees are compared. In evaluation research, a wide range of strategies have been used in order to try to overcome this problem. In what follows, we look briefly at the major approaches available in order to assess whether they could in principal be applied to the evaluation of MA.[8]

3.2 Random allocation designs

In some circumstances, random allocation research designs, otherwise known as 'social experiments', can be used to eliminate selection effects. Individuals who are eligible for the programme and (in the case of voluntary programmes) willing to enter it are randomly allocated to participant and non-participant groups. In this way, provided that numbers are large enough, all prior differences between participants and non-participants, whether observed or unobserved, should be randomised out. As a result any difference in outcomes should be attributable to the programme.

Random allocation is considered by many to be the 'gold standard' in evaluation research, and it has often been used to evaluate social programmes in North America. There is a large literature on its merits and demerits, both methodological and ethical. It is, however, pointless to review these issues here as there is no possibility of mounting a random allocation experiment to evaluate MA. With social programmes, random allocation tends to be used only where those eligible for the programme are under some form of constraint (for example, they need to draw social security benefits), where the programme is very new, or where the potential impact of non-participation is relatively trivial. The MA programme is, in contrast, well established, and it would be unthinkable to limit young people's choice of career for the purposes of research.

3.3 Instrumental variable estimation with induced variation

With this approach, we search for an 'instrumental variable' that explains variation in the probability of participating in the programme but is unrelated to the probability of achieving a successful outcome except through its effect on the probability of participation.[9] Comparing people with different values of the instrument amounts to comparing people with different probabilities of participation. Combining this with comparing participants and non-participants gives the desired estimate of the

---

[8] See Blundell 2001 for a very helpful review of non-experimental approaches applicable to this problem.
[9] The Heckman (1979) bivariate normal selection model is essentially an instrumental variable approach.

programme effect. However this effect has general limits, in that it gives the effect of the programme only for people who react to changing values of the instrument by changing their decision about whether to participate (the 'local average treatment effect'), rather than for the eligible population as a whole. A good instrument is thus one that makes it plausible to generalise the local average treatment effect to a larger population.

Unfortunately, it is often hard in evaluation research to find a valid instrument, but it is sometimes possible to create one artificially. This 'induced exogenous variation' can be thought of as random allocation at one step removed.

Consider, for example, a hypothetical training course aimed at improving the job chances of long-term unemployed people. Living in a isolated rural location may reduce both the probability of taking part in the course and the chances of getting a job, because of transport difficulties. However, if a special bus service were laid on to take people from a randomly selected set of rural districts to the course, and people were not allowed to use the bus service for any other purpose, then the chances of participation would be increased for people living in those districts, while their chances of getting a job after the programme had ended would be unaffected.

This type of evaluation strategy may sometimes be practical and ethically acceptable when direct random allocation of places on the programme is not. Indeed, it is sometimes considered to be the next best option to random allocation in terms of methodological rigour and the likelihood of producing clear results. The strategy can operate either at the local area level (as in the case of the bus service), or at the individual level (for example, if a random sub-sample of those eligible for the programme were given financial incentives to take part).

The difficulty in applying this approach to the evaluation of MA lies in finding an intervention that would increase participation in MA and still be politically and ethically acceptable if applied on a random basis. One possibility could be to increase funding for MA in randomly selected areas, with the aim of increasing the number of places available and hence participation in those areas. However, one of the reasons for creating the Learning and Skills Council (LSC) was to even out differences between local areas in funding, and so this option would be unlikely to meet with political approval. In addition, there are likely to be other constraints on the expansion of MA apart from funding levels, that have to do with the local industrial structure and the way that employers use labour. A second possibility could be to run a marketing campaign for MA in randomly selected areas where MA was under-subscribed, aimed at encouraging more young people to enter. However it is doubtful whether such a campaign would significantly increase take-up, unless it involved very strong persuasion applied to individuals, which itself would raise ethical and political

objections. A third possibility is to pay financial incentives to young people to enter MA, which the DfES may in any case be considering in the future for disadvantaged young people.[10] To apply such incentives on an individually random basis is also likely to meet with political opposition, as it did in the case of the EMA pilot studies. This last option is nevertheless perhaps the one most likely to lead to a successful research design.

3.4 Instrumental variable estimation with naturally occurring variation

With this approach, we search for an instrumental variable that occurs without any intervention from the researcher. Sometimes this may result from a 'natural experiment', as, for example, when a chance administrative problem (perhaps a local strike) prevents a new programme being set up in particular areas, or when policy varies between local areas - for example, under the former discretionary system, LEAs differed in the extent to which they used educational maintenance allowances (EMAs). In other cases, it may arise from individual level factors; for example, it has been suggested that unexpected family financial difficulties can be used as an instrument for participation in post-compulsory education.[11]

In the case of MA, it is hard to think of an individual-level factor that is related to the probability of participation in MA but is not related to later employment chances or earnings. However there is considerable variation in participation rates between areas which might be helpful in identifying an instrumental variable. YCS data suggest that amongst young people who had reached the end of compulsory full-time education in summer 1997, the proportion in AMAs the following spring varied from around 1% in Greater London to 7% in the North, while the total proportion in WBTYP for young people varied from 4% in Greater London to 21% in the North.[12] Published official statistics on young people's rates of participation in education and training by region do not distinguish between MA and OT, but amongst 16 year olds the overall rate of participation in WBTYP varied in 1997/98 from 18% in the North East to 5% in London.[13] Regional variations in participation appear to persist even when GCSE results and other individual-level factors that influence choice of route at age 16 are taken into account.[14]

---

[10] The DfES's response to the recent consultation on MA states, ' The Department will not be introducing specific financial incentives to young people to encourage completion, but may investigate whether they would be helpful in cases of market failure or to encourage disadvantaged young people into training' (see DfEE 2001).

[11] Blundell 2001.

[12] Author's computation, using Standard Statistical regions.

[13] DfEE 1999, Table 1 (Government Office Regions).

[14] Payne 1988, Chapters 5 and 6.

In order to exploit this variation in an evaluation study, we need to find an instrumental variable that is associated with local area variations in participation rates in MA but unrelated to outcomes from MA. As MAs are concentrated in particular industrial sectors, one possibility suggested in discussion was the proportion of industries in the local area with a high use of MAs. A difficulty with this suggestion is that only certain MA sectors such as Engineering Manufacturing are geographically concentrated to any significant degree. The large majority of MA trainees are in sectors that are geographically dispersed, such as Business Administration, Customer Service, Hospitality, Construction, Hairdressing, Health and Social Care, Retailing and Childcare. Another difficulty is that the local industrial structure is itself related to other local factors that affect outcomes from MA, such as the local unemployment rate and level of employment growth, which would need to be controlled for, reducing the likelihood of getting clear results.

A second possible candidate for an instrumental variable is local cultural norms regarding education and training, which may influence a young person's decision about which direction to take at age 16. These are however hard to measure and no nationally consistent data are available at a sufficiently detailed local level.

3.5 Matched area designs

The DfES and the ES have often used a matched area design to try to assess the impact of a new programme. With this design, the programme is launched in a small number of pilot areas, and a number of other areas thought to be comparable in relevant respects are selected in order to compare performance on the outcomes that the programme is designed to affect. This design is currently being used to test whether Employment Zones reduce rates of long-term unemployment, and whether EMAs increase the rate of participation in post-compulsory full-time education.

Matched area designs have a number of problems, but it is irrelevant to discuss these here. There is no possibility of using this strategy to evaluate MA, as the programme has been running nationally for several years and cannot be selectively withdrawn.[15]

3.6 Difference in differences

When experimental evaluation designs are ruled out, the counterfactual question of what would have happened to programme participants if they had not gone on the

---

[15] It has been suggested that Scotland might provide suitable comparison areas in a matched area design. However the differences in the education and training systems between England and Wales on the one hand and Scotland on the other are probably too extensive to make this a viable strategy.

programme can sometimes be addressed by the 'difference in differences' strategy. This approach was used in the evaluation of the New Deal for Young People (NDYP).[16] The rate of leaving unemployment amongst long-term unemployed people in the NDYP age group was compared with the rate of leaving unemployment amongst an older age group of long-term unemployed people, both in the period following the launch of the programme and also over a period that predated its launch. Provided that we can assume that there were no other changes over time that affected the difference in leaving rates between the two age groups, we can attribute any change over time in the difference between their leaving rates (adjusted for any changes in their composition) to the impact of NDYP. In this design, selection effects are automatically dealt with, in that they are assumed to be constant in the before and after periods, and so cancel out.

In order to apply this strategy to the evaluation of MA, there would need to be a clear target group at whom the programme was aimed, and a similarly clearly defined group who were not eligible for the programme with whom those eligible for MA could be compared. However MA is open to all young people aged 16-24, with no formal restrictions on entry, so we cannot identify target and non-target groups within the 16-24 age group. Nor would it be plausible to make comparisons with older age groups outside the eligible age range, given the comparatively small proportion of 16-24 year olds involved in MA, the importance of other policy changes that have affected the job chances of 16-24 year olds (notably NDYP), and the other factors that are likely to have changed the relationship between the wages of younger and older workers (for example, the minimum wage legislation). A further problem arises from the fact that, unlike NDYP, the launch of MA did not mark a distinct change in policy: AMAs developed out of the apprenticeship training that had been available for many years within government-supported training for young people, and FMAs are part of a steady progress over a number of years towards improved standards of training and a greater stress on working towards nationally recognised qualifications.

3.7 Matched comparison group designs

A matched comparison group evaluation design deals with the problem of selection effects by comparing the progress of programme participants with that of eligible non-participants who are similar to participants on observable factors that affect both the chances of participation and the chances of a successful outcome. The design is in principal not as strong as a random allocation design, in which all unobservable as well as observable differences between participants and non-participants are randomised out. By definition, a matched comparison group can only be selected on

---

[16] Wilkinson 2001.

the basis of observable characteristics, and it is always possible that, however careful the matching, there will remain unobserved differences between the participants and comparison group that contribute to any apparent programme effect. This problem can be mitigated by matching on as wide a range of variables as possible, and by including variables that can proxy for unobservable factors. For example, in an evaluation of a training programme for unemployed people, previous experience of unemployment could partially capture unobserved differences between participants and non-participants in motivation, self-confidence and self-presentation.

The rationale of the matched comparison group design is the same, whether the comparison sample is located and surveyed especially for the evaluation study, or whether it is extracted from an existing data set such as the Labour Force Survey (LFS) or YCS. The main obstacle to a new survey is often that of finding a suitable sampling frame for non-participants, and further difficulties can arise because sample attrition is likely to lead to a proportion of the original matches being lost. The main problem in using an existing data set is often that it does not contain all the variables that are needed in the analysis.

A range of sophisticated matching techniques is now available under the general head of 'propensity score matching', which allow a large number of variables to be used in the matching process. Propensity score matching has other statistical advantages that make estimates of programme effects more accurate. In addition, programme effects can be reported in terms of simple averages or percentage point differences between the participant and comparison groups. These are much easier to present to a non-specialist audience than the results of a statistical model.

The choice of matching technique to construct the comparison group will depend to a large extent on the number of variables available in the frame from which the sample is drawn and the size of the pool from which the matches are drawn. If there is only a limited number of variables (as, for example, in JUVOS), then a simple matching technique may have to be used, and a wider range of potential control variables to use when estimating programme effects would have to be collected in a subsequent survey.

Whichever matching technique is used, the viability and plausibility of a matched comparison group approach depends on the nature of the selection process into the programme, and hence the strength of the selection effects. Problems arise if the selection process differentiates very strongly between participants and non-participants, and this can also apply when programme entrants are self-selected. There are also likely to be problems if participants form either the overwhelming majority or only a tiny proportion of those eligible for the programme. In these cases,

it may be impossible to find enough non-participants who match participants reasonably closely on observed characteristics (the 'support' problem).

In the case of MA, selection into the programme does not appear to be so strong as to make a matched comparison design unsuitable. Everyone aged 16-24 is eligible for MA, and there are no formal entry requirements for the programme as a whole.[17] In 1999 around 9% of all 16-18 year olds were in WBTYP, of whom the majority were in MA.[18] The dispersion in post-16 activities is much wider amongst those with relatively poor GCSE results than amongst those with good results. In spring 1998, 18% of 16/17 year olds with below-average Year 11 GCSE results were in WBTYP including 6% in MA. This compares to 16% in full-time jobs outside of WBTYP, 49% in full-time education and 17% with no full-time activity.[19] It seems likely that in addition to the systematic influences on young people's choices at 16, chance factors also play quite a big role, such as the particular opportunities that crop up, the encounters that young people have with others or personal and family events that necessitate changes of plan.[20] However both AMA and FMA are concentrated within a relatively narrow range of sectors compared to the youth labour market as a whole, which creates complications when trying to find an appropriate comparison group for MA trainees in certain occupations.

The main difficulty in applying the matched comparison group strategy to the evaluation of MA lies in finding a suitable data set from which to construct the matched groups, or a suitable sampling frame from which to identify a comparison sample for a new survey. These issues are discussed in Section 4.2 and in Section 5.

3.8 Statistical models without prior matching

In the least sophisticated type of evaluation design, a statistical model of the outcome of interest is fitted and all relevant variables in the data set are included as predictors along with a variable indicating programme participation or non-participation, with no prior matching between these groups. Like propensity score matching, this method assumes that all prior differences between programme participants and non-participants are observable. The approach is only likely to be used if a data set is

---

[17] Proposed minimum entry requirements for a place on MA are intended to be 'a guide for applicants rather than a restrictive measure' (DfEE draft consultation document on MA, Feb 19 2001), and such requirements are in any case likely to be set quite low.

[18] DfEE 2000a (provisional figures). Estimates of the proportion of the age group in MA specifically are not published.

[19] Author's computations. Having 'below average GCSE results' is defined as being in the bottom half of the national distribution of Year 11 GCSE results, based on total GCSE points score. The figures are derived from Sweep 1 of YCS Cohort 9, who reached the end of compulsory education in summer 1997.

[20] Examples of all these can be found in DfEE 2000c.

already available for analysis - if a new survey is needed, then it will generally be more efficient to collect data on a sample of non-participants that is matched to participants in some way. As a method of analysis, it is probably inferior either to propensity score matching or to techniques that attempt to adjust estimates of programme impacts for the effects of unobserved differences between participants and non-participants. The key difference between matching and just running a regression on the same variables is that matching relaxes the linear functional form restriction implicit in the regression, thus increasing the likelihood of detecting real relationships in the population from which the sample is drawn.

3.9 Conclusion

This review of the main strategies that can be used to measure the net impact of social programmes suggests that most are not viable in the case of MA. Possibly the strongest design would be an experiment in which financial incentives to enter MA were randomly allocated, but it is questionable whether political agreement could be secured for this. Otherwise the matched comparison group design would probably have the lowest bias amongst those design options that are politically and practically feasible, provided that data of sufficient quality could be obtained. In Section 5 we discuss a number of data sets that could be thought of as candidates for this role. First, however, we consider some of the issues that need to be dealt with in developing an evaluation design using the matched comparison group methodology.

4 SPECIFIC ISSUES IN EVALUATING MA

Section 3 concluded that a matched comparison group design offers the best hope among the politically and practically feasible estimators of measuring the net impact of MA on trainees' job chances and earnings. In this section we consider a number of issues that need to be dealt with in developing an evaluation using this methodology. The choices to be made should be guided by policy interests, but will also be heavily constrained by practicalities.

4.1 Choice of comparison group - policy issues

Any estimate of a programme's net impact on individuals is relative to a counterfactual case, namely what would have happened to them if they had not gone on the programme. In some cases it is obvious how this counterfactual should be specified - for example, what happens to people who take part in a new training programme for the unemployed can be compared with what would have happened to them if they had received only the standard services for the unemployed. With MA trainees, the appropriate counterfactual is not so clear, as there are a number of different routes that they could have taken instead. Estimates of the net impact of MA will vary according to whether it is compared with full-time study for vocational qualifications, with jobs outside of WBTYP, with OT, or with unemployment or economic inactivity. All these comparisons have some interest.

If it is not possible to extract a comparison sample from an existing data set and the comparison sample has to be identified and surveyed especially for the MA evaluation study, then deciding on the appropriate counterfactual becomes particularly difficult. Taking more than one comparison group would be very expensive, and even if resources permitted two or more comparison groups, each one would involve a different sampling frame, with different biases in coverage. These difficulties would probably limit the study to one comparison group only. As the ultimate purpose of the evaluation study is presumably to assess the value of expenditure on MA compared to expenditure on other policy options, the choice of which comparison group to take would need to be based on the most dominant policy interests.

Discussions have suggested to us that there is not much interest within the DfES in comparing MA trainees with students in full-time education, whether on academic or vocational courses. This is because the choice between work-based training and full-time education is seen to depend on young people's individual preferences, and it is felt to be important in principle to offer them the choice between these two routes.

Nevertheless, there has been a substantial shift since the 1980s away from work-based training for young people towards vocational courses in full-time education. This trend clearly does not result purely from changes in individual preferences, but reflects wider economic and policy developments which the introduction of mandatory EMAs may reinforce. It therefore seems useful to compare outcomes for young people in MA with outcomes for those on the other major route by which vocational qualifications can be gained.

The DfES has also indicated that there is little point in comparing MA with OT, as this option is due shortly to be phased out. Nevertheless this comparison could be of more general policy interest as it could assess the value of the reforms to government-supported training for young people embodied in MA. More specifically, it could give some indication of the returns to better structured, better quality training.

There could also to be value in comparing MA trainees with young people in jobs outside of WBTYP, as this could indicate the returns to government investment in subsidised and well-structured youth training compared to an unregulated system in which all the costs of any training that is carried out are borne by employers or employees. There may be particular interest in the sectors where the MA framework has only recently been established or has not yet gained much of a foothold, such as the high technology industries, retailing and the cultural heritage industry.

There may also be interest in comparing MA trainees with young people who are not in education, employment or training (NEET). The present government has expressed considerable concern about socially excluded young people, and one of the aims of the reforms to MA currently being planned is to encourage more of this group to enter the programme.[21] However interpreting a comparison between MA trainees and the NEET group would require care, as part of the impact for them of entering MA would stem simply from the move from unemployment or economic inactivity to the workplace, rather than from the specific training provided.[22]

The problem of choosing an appropriate comparison group is much easier to resolve if, instead of mounting a special survey to collect information on the comparison group, data can be extracted from an existing or planned large data set that has information on a representative sample of all young people in the relevant age group. This would free the evaluation study from budgetary pressures to choose a single comparison group, and different comparison groups could be selected to address different issues. Some possible data sets that could be used for this purpose are discussed in Section 5 below.

---

[21] See Social Exclusion Unit 1999 for a discussion of these issues.
[22] We are grateful to a participant in the April 5 2001 workshop for this point.

4.2 Choice of comparison group - practical issues

If there were to be a new survey of a comparison group or groups (rather than extracting these from an existing large data set) we would need to have sampling frames for these groups. Without a sampling frame, the only alternative is a prior population screening survey, which would be very expensive.

- For those in OT the Trainee Database System (TDS) could be used. The viability of this as a sampling frame is discussed in Section 6.

- At the moment young people in full-time vocational education can only be identified via schools and colleges. Co-operation in the sampling process is very burdensome for these institutions and quite likely to be withheld. The 1998 survey of MA trainees and a comparison sample of young people on A level and GNVQ courses achieved responses from only 119 of the 200 schools and colleges that it sampled, and there is no reason to suppose that a new study would be any more successful.[23] This problem may be solved when the planned Individualised Pupil Record (IPR) comes into operation nationally next year, as this could be used as a sampling frame.[24]

- Young people in jobs outside of WBTYP would have to be identified via their employers, for whom a suitable sampling frame exists in the form of the Inter-Departmental Business Register. There is also a possibility of identifying them through the Learning and Training at Work Survey, an annual survey of around 4,000 employers in England with one or more employees, which includes questions on participation in WBTYP. In either case, a big burden would be imposed on employers, who would have to be asked to supply names of individual employees, and the level of co-operation is likely to be low.[25]

- Young people not in education, employment or training tend not to appear at all in official records until they reach 18, when they may claim unemployment benefit. This situation may improve as better administrative systems are developed to track young people between full-time education, the youth services and work-based training, but at the moment there is still a long way to go before such

---

[23] Coleman and Williams 1998.

[24] The DfES is already planning to use the IPR as the sampling frame for the proposed new longitudinal survey of young people (see Section 5.3.1).

[25] The 1998 Workplace Employee Relations Survey (WERS) had a response rate of 80.4% amongst sampled workplaces that proved to be in scope (i.e. still in operation and with more than 10 employees). Of these, 86% agreed to provide a fixed-size sample of employees. The response rate in this employee sample was 64%. This suggests an overall response in the employee survey of 44%. WERS covers topics of great relevance to employers and is regarded as exceptionally successful in obtaining their co-operation; the response from employers asked simply to provide a sample of young workers is likely to be much poorer. See Cully *et al.* (1999).

systems can be put in place. Another difficulty lies in the fact that young people who are NEET at a given point in time tend not to remain NEET long-term, but to move in and out of NEET status.[26] Thus they are best defined not by their status at any given moment, but by their history over a period of months or years, which makes the task of identifying them more difficult.

If the comparison sample is drawn from an existing large data set rather than a new survey, there are still likely to be problems with obtaining a satisfactory sample of young people not in education, employment or training. The NEET group tends to be under-represented in population surveys, for two reasons. First, they may not appear in the original sampling frame for the survey (for example, they may be excluded from school and so not appear in a sample based on school enrolments, or they may be homeless, and so not appear in a household survey). Second, they tend to have a lower response rate than other young people.[27] In addition, although there may be a reasonable degree of overlap between MA trainees and the NEET group on easily measured characteristics such as GCSE results and parental occupation, other factors that are hard to quantify in surveys may be important. For example, the NEET group may be more heavily involved in drugs, alcohol and crime.

One difficulty in comparing MA trainees with young people in jobs outside of WBTYP, arising regardless of whether the comparison sample is drawn from an existing data set or from a new survey, is that in some jobs, such as vehicle trades and hairdressing, WBTYP accounts for nearly all young workers.[28] For these occupations, it would be difficult to separate the effects of doing an MA from the effects of choosing that occupation rather than another. This is not a problem if we focus on occupations where MA accounts for a smaller share of young workers, and we have already suggested that a focus on these occupations may be of more policy interest.

4.3 Age at which outcomes are measured

The best age at which to measure outcomes from MA depends on the usual age of entry to training, which varies considerably between AMA, FMA and OT. As Table

---

[26] Payne 2000.

[27] An indication of the extent of this problem is given by the fact that in Sweep 1 of YCS Cohort 9, young people who were unemployed are given a mean weight of 1.28 to correct for differential non-response. For some subgroups, for example low achieving males, mean weights are much bigger.

[28] At Sweep 1 of Cohort 8 of the England and Wales Youth Cohort Study, 89% of 16/17 year olds working in SOC Minor Group 54 (Vehicle Trades) were in WBTYP, as were 87% of those working in SOC Minor Group 66 (Hairdressers, Beauticians and Related Occupations). Other occupations where WBTYP was the predominant mode of entry were SOC Minor Groups 52 (Electrical/Electronic Trades) (87%) and 57 (Woodworking Trades) (91%). (Source: Special analysis conducted for this report by Joan Payne.)

TABLE 1
Age on entry to AMA, FMA and OT, by financial year (England and Wales)

		1996/97 %	1997/98 %	1998/99 %	1999/2000 %
AMA					
	16 years	19	20	19	17
	17 years	18	19	18	17
	18 years	18	18	19	18
	19 years	13	12	13	14
	20 years	9	9	9	10
	21 years	8	7	7	8
	22 years	7	7	6	7
	23 years	5	5	5	6
	24 years	3	2	2	3
Total		*100*	*100*	*100*	*100*
N		*75,241*	*87,713*	*85,865*	*87,255*
FMA					
	16 years		27	37	33
	17 years		34	27	24
	18 years		16	15	14
	19 years		6	7	9
	20 years		5	4	7
	21 years		3	3	5
	22 years		4	3	4
	23 years		3	2	3
	24 years		2	1	2
Total			*100*	*100*	*100*
N			*1,398*	*41,619*	*99,103*
OT					
	16 years	37	41	42	41
	17 years	24	28	31	33
	18 years	9	9	10	9
	19 years	7	6	5	5
	20 years	5	4	3	4
	21 years	5	4	3	3
	22 years	5	3	2	2
	23 years	4	3	2	2
	24 years	3	2	1	2
Total		*100*	*100*	*100*	*100*
N		*255,813*	*204,271*	*130,490*	*84,842*

Source: Special analysis of the Trainee Database System conducted for the feasibility study.

1 shows, in 1999/2000, only around one in six entrants to AMA were aged 16, compared to one in three FMA entrants and two in five entrants to OT.[29] Around one in three AMA entrants were aged 20 or more, compared to one in five entrants to FMA and one in eight entrants to OT. These older entrants will not expect to complete their training until their mid-twenties.

---

[29] We are grateful to James Geehan of the DfES for supplying these figures.

TABLE 2
Proportion of all AMA and FMA starts that are at age 16, by sector:
England and Wales, financial year 1999/2000

	% of starts at age 16	Base N
AMA (11 largest sectors)		
Business Administration	16	10,235
Early Years Care and Education	8	3,378
Electro-technical Industry	40	3,489
Engineering Manufacturing	37	7,217
Retailing	3	5,496
Construction	24	4,754
Hairdressing	27	5,318
Hospitality	8	7,987
Motor Industry	39	5,990
Health and Social Care	3	5,738
Customer Service	3	8,524
All other sectors	16	16,794
FMA (11 largest sectors)		
Business Administration	43	15,405
Early Years Care and Education	39	1,669
Engineering Manufacturing	31	5,310
Information Technology	46	5,201
Retailing	21	13,961
Construction	48	6,170
Hairdressing	63	8,176
Hospitality	16	10,338
Motor Industry	55	3,807
Health and Social Care	17	4,653
Customer Service	10	8,663
All other sectors	30	12,068

Source: Special analysis of the Trainee Database System conducted for the feasibility study.
Note: OT is not organised by sector.

The situation is complicated by the fact that the average age of entry to MA differs between sectors. As we see in Table 1, the average age of MA entrants has risen as MA has expanded into new sectors.[30] Table 2 shows that while around two in five AMA starts in traditional craft apprenticeships like Engineering Manufacturing, the Electro-technical Industry and the Motor Industry were at age 16, for the newer apprenticeship sectors like Early Years Care and Education, Retailing, Hospitality, Health and Social Care and Customer Service the corresponding proportion was less than one in ten, and sometimes less than one in twenty. FMA trainees are on average younger on entry than AMA trainees, but as Table 2 also shows, there is a similar pattern of variation between sectors. Thus while 19 or 20 might be an appropriate age

---

[30] Average age on entry was higher in the first year of AMA and FMA than in the second year because a number of young people transferred to MA from other WBTYP programmes.

at which to measure outcomes from MA in traditional sectors, this would be far too early for many of the newer sectors.[31]

The wide spread of ages on entry to MA means that using existing cohort studies to evaluate MA would give only very partial coverage of the programme. This issue is discussed in more detail in Section 5.

The age at which outcomes should be measured will also depend on the choice of comparison group. If comparing MA trainees with those on OT, young workers in jobs outside the MA framework or the NEET group, then it could be acceptable to measure outcomes soon after the completion of MA training. This could also be acceptable if comparing MA trainees with those who stay in full-time education for only one or two years after the minimum leaving age. However for comparisons with young people who enter higher education, the earliest age at which outcomes can be sensibly measured is probably 24 or 25, given that many university students now have a gap year before entry or take time out after completing their degrees.

Clearly the older the age at which outcomes are measured, the better the indication will be of long-term effects. However, if the evaluation study is prospective, then the older the age, the longer the wait for results and the greater the problems of sample attrition. These problems are avoided if the study is retrospective rather than prospective in design, though here constraints are imposed by the date that MA was launched.

4.4 Prospective versus retrospective designs

With a prospective evaluation design, a sample of young people would be identified at or before entry to MA, and then followed over a number of years. With a retrospective design, a sample of people who had been on MA in the past would be identified, who would then be asked to recall information about what they did. A prospective design is in theory much stronger than a retrospective study, for several reasons.

- People's recall is often very faulty, particularly where government programmes are involved.
- Information on some variables that are important predictors of programme participation or of outcomes, such as attitudes, cannot be collected retrospectively.
- It is much easier to track movements into and out of MA and transfers within the programme prospectively than to disentangle them afterwards.

---

[31] There may be a reduction over time in the average age of MA starts in the newer sectors as the existing stock of young people interested in training in these fields is diminished and future trainees are drawn only from the flow of young people who age into eligibility.

The main difficulty with a prospective design is that it would take at least four years to produce results, whereas the timescale for a retrospective study could be quite short. In addition, some of the problems involved in collecting data on participation in MA retrospectively from former trainees can be avoided by tracing them in administrative records. This is possible because the TDS stores information at the individual level on all WBTYP entrants (see Section 5.1.3).

4.5 Entry cohorts, leaving cohorts and age cohorts

While a prospective study would be based on an MA entry cohort, a new retrospective study could be based on a cohort of either entrants or leavers, as both can be identified in the TDS. The DfES's regular postal follow-up surveys of WBTYP participants are based on leaving cohorts, and give information on employment and earnings six months after leaving. This is entirely appropriate for the purpose of monitoring programme outcomes. However for evaluation purposes it is more useful to take an entry cohort, as this allows us to have a common date for entry to the programme and for measuring outcomes, with a standard interval between during which uniform conditions in the national economy prevail. The difficulty of tracing trainees is of course greater for an entry cohort than for a leaving cohort, an issue which is discussed in more detail in Section 6.6.

Section 2 reviewed some of the developments in the MA system since its launch and suggested that the very early entry cohorts may not necessarily give a good guide to later performance. For AMA, the earliest entry cohort that could be used in an evaluation study is probably autumn 1997, but there were very few entrants to FMA that year. This issue is discussed further in Section 6.3.

Another alternative is to base the evaluation study on an age cohort of young people, which, given the wide age-range of MA entrants, would entail a wide range of calendar dates for both entering and leaving MA. If this involved a new longitudinal study it would mean a very long wait for results. A serious limitation of basing the evaluation on existing age cohort studies with a relatively short follow-up period (such as YCS) is that all who were aged more than 16 or just turned 17 on entry to MA would have to be excluded from the analysis, as their training would not be completed by the date of the last follow-up survey.

4.6 Drop-out

The drop-out rate from MA is quite high: in 1999-2000 one third of leavers from AMA had not gained any full qualification and more than half had not gained any full qualification at Level 3 or above.[32] YCS data for young people in AMA in the spring following the end of compulsory education in 1998 suggest that about 30% left WBTYP altogether over the following 12 months.[33] It is important for the evaluation study to measure the level of drop-out more precisely and to estimate the net impact of MA on the job chances and earnings of young people who do not complete their training. We know from YCS data that young people's pay in AMA and OT is lower than in jobs outside WBTYP, and that trainees who leave AMA or OT early in order to take a full-time job tend to improve their pay in the short-term.[34] The interesting issue is whether early leavers maintain this advantage in the long-term, after those who stay on in MA have completed their training. This question is more easily addressed with an entry cohort followed for a fixed period of time than with a cohort of leavers.

4.7 Estimates for AMA and FMA

We understand that the reforms to MA that the DfES is currently planning are likely to mean that in future there may not be such a clear distinction between AMA and FMA as at present. Instead there is likely to be more of a 'pick and mix' approach to MA's proposed three elements (NVQs, Key Skills and the Technical Certificate). Already a number of young people transfer in both directions between FMA and AMA. Despite this, we still think that the evaluation study should be capable of providing separate estimates of the net effects of AMA and FMA, firstly because this will give information on the value of training to different levels, and secondly, because this would be helpful in estimating the effects of the MA programme on the wider economy. Even if separate estimates for AMA and FMA were not needed, trainees from both programmes should be represented in the sample on which the evaluation is based. This requirement has implications for our choice of data set and cohort.

4.8 Geographically-based information

It will be important for the evaluation study to include geographical information that enables data on the local economy to be linked to individual sample members. This is

---

[32] DfEE 2000b, Table 3.
[33] Payne 2001, Table 4.2.
[34] Payne 2001, Tables 4.8 and 4.10.

needed both to estimate the impact of MA at the individual level and for the macro-analysis discussed in Section 7. Ideally this information should be obtained both at the point of entry to training and at the time at which outcomes are measured.

5 POSSIBLE DATA SOURCES

In this section, we review a number of data sources that appear at first sight to be suitable for use in an evaluation study of MA. In some cases the data contain a sufficiently rich range of variables to be suitable for the matched comparison group methodology; in others, only a more limited analysis would be possible. The adequacy of the information on the two main outcome variables, earnings and employment rates, also varies. Overall, the data sources that have the richest range of variables and the best possibilities for identifying appropriate comparison groups unfortunately tend to have the smallest sample numbers for MA trainees.

5.1 Data already available for analysis

5.1.1 YCS Cohort 9

At first sight, YCS has some attractions for the evaluation of MA. It covers a complete age group of young people with a large sample size (around 15,000 at Sweep 1), so offers the possibility of defining more than one comparison group. It has a nationally representative unclustered sample that can be weighted to correct for non-response. There is a longitudinal structure, with a well-developed set of questions on economic activity, work-based training and post-16 qualifications, enabling us to track progress through MA and to identify entrants at different ages and those who drop-out from training. There is also a reasonable set of variables to control for selection effects, including full information on GCSE results allowing a GCSE total points score to be calculated, some details of home background, some school variables, careers advice and guidance in Year 11, and attitudes to education, work and training, though all information relating to the period before the end of compulsory education is collected retrospectively. Among the weaknesses of YCS is its heavy reliance on postal questionnaires with a consequent response bias against low attainers (although response is now boosted by telephone interviews with non-respondents). There are also high levels of sample attrition across sweeps.

YCS Cohort 9 is the first YCS cohort that could be considered for the MA evaluation study. The previous cohort, YCS Cohort 8, reached the end of compulsory full-time education in summer 1995, the year that AMA began. FMA did not yet exist. Sweep 1 of Cohort 8 took place in spring 1996 at age 16/17, Sweep 2 in spring 1998 at age 18/19, and there was a third sweep in early autumn 2000 at age 21/22. At Sweep 1 294 respondents were in AMA, of whom 150 responded at Sweep 2 (unweighted sample numbers). Data for age 21/22 are not yet available, but the response rate was

around 58% of Sweep 2 respondents. This will probably yield a sample of around 85-90 Cohort 8 members who were in AMA at 16/17, which is unlikely to be large enough to detect any programme impact.

YCS Cohort 9 reached school leaving age in summer 1997, after the launch of FMA (then known as NTr), and when numbers on AMA had grown substantially. Sweep 1 took place in spring 1998, Sweep 2 in spring 1999 and Sweep 3 in spring 2000. Unusually for YCS, there was a fourth sweep in autumn 2000, when respondents were aged 19 or only just 20; data from this are expected to be available in October 2001. The cumulative effect of sample attrition across sweeps is to reduce Sweep 4 sample numbers to around 31% of the Sweep 1 figure.

Although there is no fixed duration for AMA, a period of around three years to completion is normal. It follows that Cohort 9 could be used to estimate the impact of AMA for young people who were in AMA at Sweep 1, but could not be used to estimate impacts for older entrants, as many of these would still be in training at Sweep 4. Given that in 1999/2000 only 17% of AMA trainees were aged 16 on entry and that the average age on entry differs between sectors (see Section 4.3), this would limit the generalisability of the results.

Allowing for attrition, there may be information at Sweep 4 for around 170 of the 544 cohort members (unweighted numbers) who were in AMA at Sweep 1. This may be just enough to detect a programme impact provided that the impact is fairly large.[35] However the sample would be much too small to permit any analysis of subgroups, or to detect any differential effects for different sectors within AMA, for different types of trainees or for different types of training provision such as the amount of off-the-job training.

For FMA sample numbers are much too small for any analysis. Only 49 respondents were in FMA at Sweep 1, of whom perhaps 15 have data at Sweep 4. There is little point in combining these with the AMA sample as they would only confuse the interpretation of results.

Another serious limitation of an evaluation study based on YCS Cohort 9 is that it would be restricted to the impact of AMA on job chances, as questions on earnings

---

[35] If we were comparing a simple random sample of 170 AMA trainees with a simple random sample of non-participants of equal size, and around 70% of AMA trainees found work, then there would need to be a difference of around nine percentage points in the proportion in work between the two samples for the difference between them to reach significance at the 95% level of confidence. Increasing the size of the non-participant sample changes this only slightly. This calculation takes no account of design effects, which would tend to increase the size of the required difference between the two samples.

were omitted from Sweep 4. [36] Although information on earnings was collected at Sweep 3, many 16 year old entrants to AMA would still be in training at that point, and their earnings while in AMA cannot be taken as a guide to their earnings after completion.

5.1.2   The Labour Force Survey

The LFS sample covers around 60,000 domestic addresses and around 160,000 individuals in any one quarter. All people aged 16 or over living at the address are surveyed, so there is good coverage of the MA age group. The survey has a panel structure, with one fifth of addresses replaced in each quarter, so that each stays in the sample for five interviews. This means that the first and last interviews are just 12 months apart. In the case of an evaluation of MA, this is probably too short a period to make it worthwhile to exploit the panel structure by tracking those in MA at their first interview across their next four interviews. Instead the LFS is probably best used cross-sectionally to compare people who have already completed MA training with others.

The main LFS question on apprenticeship is, *'Are you doing, or have you completed, a recognised trade apprenticeship?'* This form of words is usually associated with skilled manual occupations and may lead to an under-counting of trainees in the newer MA sectors such as retailing or the cultural heritage industry. The problem is likely to be made worse by the LFS's extensive use of proxy interviews. The extent of any possible under-counting of MA trainees in the newer sectors could be estimated by comparing LFS estimates of the numbers in MA training in different occupations with figures in the TDS, though there may be difficulties in some cases in mapping MA sectors onto the Standard Occupational Codes used in the LFS.

The coded responses to the main question on apprenticeship are:
> *1 yes (completed)*
> *2 yes (still doing)*
> *3 no (including apprenticeship begun but discontinued).*

This means that the LFS cannot be used to compare outcomes for those who completed their apprenticeship and those who dropped out before completion. Given the high level of drop-out from MA, this would be a serious problem for an evaluation study.

---

[36] The reason for the omission was that Sweep 4 was primarily designed to provide information on progress towards the national attainment targets, and it was felt that including a question on earnings would lower the response rate.

The job title of the apprenticeship is only asked of those who are still doing their apprenticeship at the time of survey. As a result it would not be possible to use the LFS to compare outcomes from training in different MA sectors.

Modern Apprentices are identified by the following question asked of those who have completed or are still doing a recognised trade apprenticeship: *'Does/did your apprenticeship form part of the Modern Apprenticeship initiative?'* Responses are coded as follows:

> *1 yes*
> *2 no*
> *3 never heard of Modern Apprenticeships.*

Thus the LFS cannot distinguish between AMA and FMA. FMA was known as National Traineeships (NTr) up until 2000, so most LFS respondents who say they have completed an apprenticeship that was part of the MA initiative can be assumed to have been in AMA. A separate section of the questionnaire covers Youth Training, but does not refer to MA or NTr.

The LFS has no information on the date that the apprenticeship was completed or on age at completion.

In LFS data relating to the summer quarter 2000, 12,552 respondents said that they had completed a recognised trade apprenticeship, of whom 653 said that their apprenticeship was part of the MA initiative. This number could be boosted by aggregating data across quarters, omitting repeat observations for individuals already included in the sample. In a full year's LFS sample, there may be around 1,000 respondents who have completed MA.

The LFS has very good data on the two outcome variables in the evaluation, namely employment and earnings, but has only a limited set of variables that could be used to control for the effects of selection into MA. These include age, sex, ethnicity, age of leaving continuous full-time education, qualifications gained in school (proxied) and possibly health and disability. The main problem for evaluation purposes is the limited amount of historical data: there is information on qualifications, household circumstances and health/disability at the time of survey, but no information that can be securely dated to the period before entry to MA.

Results in GCSEs, which are mostly taken in school, could be an adequate measure of qualifications before entry to MA. However information on these is limited, and it is possible to distinguish only the following groups: no GCSE passes, GCSE passes below grade C only, 1-4 GCSE passes at grades A-C, 5+ GCSE passes at grades A-C, with the last two groups sub-divided according to whether the respondent also has GCSE passes below grade C. Given that results in Year 11 GCSEs is by far the most

important predictor of route and progress after age 16, this does not offer a very satisfactory control for selection effects into MA. The difficulty is increased by the fact that young people who take work-based routes tend to have below average GCSE results, so that the LFS offers even less discrimination amongst the group of most relevance to the evaluation study.[37]

Because information is collected on everyone living at the sampled address, it is possible, by linking records within households, to get information on home background if young people are still living in the parental household. However young people who still live at home in their early twenties differ in many ways from young people who have already left the parental home, and excluding the latter from the evaluation is likely to give very biased results.

In addition to information on disabilities or health problems at the time of survey, the LFS also has information on past health problems or disabilities that lasted longer than one year. Although these problems would not necessarily pre-date MA entry, past health/disability might be used as a proxy for health/disability before MA entry, given that everyone who has completed MA must still be quite young.

The LFS has data on how long respondents have lived at their present address and on previous area of residence if they have lived at that address for less than 12 months. However, it is only possible to establish area of residence before entry to MA for those who have lived at their present address for a number of years, who are of course a biased sub-sample of the MA group.

The choice of comparison groups in the LFS is restricted by the lack of information on education, employment and training history from age 16 onwards. To estimate differences between the job chances of people who have completed MA and others, they could be compared with people of a similar age who have either not completed or not entered the programme, with perhaps separate comparison groups defined by age of leaving full-time education and sex. To estimate differences in earnings, the comparison would need to be restricted to people in work at the time of survey. The LFS has of course very rich information on current job, and a wide range of control variables for earnings could be applied. However these comparisons could not be said to provide an unbiased estimate of the net impact of MA, as they will be heavily affected by unobserved differences between the MA and comparison group.

---

[37] In YCS Cohort 9, nearly half of young people in AMA in the spring following the end of compulsory education were in the bottom third of the national distribution of GCSE results based on total points score, with less than a tenth in the top third. Those on OT and in jobs outside WBTYP tended to have even poorer results (Payne 2001, Chart 2.2).

Although the very limited nature of LFS information on MA makes it not particularly promising on its own for evaluating MA, it could be used to provide comparative data for information on the earnings and job chances of MA trainees obtained from other sources. This possibility is discussed in Section 5.1.3 in relation to the TDS and in Section 5.3.4 in relation to a possible new survey of a cohort of MA entrants identified from the TDS.

5.1.3 The Trainee Database System

The TDS covers everyone who starts WBTYP, with separate records for AMA, FMA, OT and the 'Life Skills' programme introduced in 1999 to help young people who are not yet ready to enter MA.[38]   It holds data derived from both administrative sources and a follow-up survey. With 142,400 young people in AMA and 95,800 in FMA in England and Wales in September 2000, sample numbers are enormous. The DfES managed the TDS, but in April 2001 responsibility passed to the Learning and Skills Council.

The administrative data in the TDS recorded at the start of training includes start date, whether has employed status, expected duration of training, planned training occupation, level and subject of qualification sought, sector, whether in guarantee group, training provider and TEC. When the young person leaves, information is added on leaving date, date of transfer to employed status if applicable, immediate destination on leaving so far as this is known, and qualifications gained in training. However there is much missing information on the last of these, as many trainees are not awarded their qualifications until after they have left WBTYP.[39]   Administratively collected variables that could be used to control for selection effects are limited to date of birth, sex, whether has a limiting health problem or disability, ethnicity, special training needs, previous participation in WBTYP, and whether in work or on Learning Gateway or New Deal Gateway immediately before entering WBTYP.

Information is also added from a postal survey six months after leaving which, despite having a very well designed questionnaire, has a response rate of well under 30%. The survey is focussed mainly on getting information on outcomes, and covers whether the training was completed and agreed objectives met, views on the training, qualifications gained, Key Skill Units completed, present activity, details of current job including earnings, relationship to the training received in WBTYP and training provided, and any qualifications now being sought. Information on hours worked is not collected, so hourly earnings cannot be calculated, though part-timers can be

---

[38] Known as 'Skill Build' in Wales.

[39] Training providers are asked to supply information on late awards in the 'outcomes collection round', which can continue for up to 18 months after the leaving date, but they have no particular incentives to do this.

identified. The only information collected in the survey that could be used to control for selection effects is on activity status immediately before starting WBTYP and qualifications before entry. Information on GCSEs is collected in much greater detail than in the LFS, and though it is not possible to calculate a total points score from the information, it probably captures enough of the variation in results amongst WBTYP trainees to function well as a control variable.

Using the administrative data to make some correction for differential non-response to the postal survey and the limited range of control variables available, the TDS could be used to compare outcomes between AMA, FMA and OT. However the value of these comparisons is limited, as the follow-up information relates to varying intervals from entry, depending on the programme and whether the trainee left before completion. Nor could these comparisons be interpreted as indicating the net impact of AMA or FMA relative to OT, because selection effects between the programmes are strong, and unlikely to be adequately captured by the variables available in the TDS.[40]

Individuals in the TDS are identified by their NINO, so, provided that data protection requirements could be satisfied, it is in principle possible to link the TDS with information from JUVOS or from the New Earnings Survey (NES). This opens the possibility of longer-term and more complete information on outcomes than is currently available. JUVOS records all claims for unemployment-related benefits, and would provide information on trainees' future spells of claimant unemployment. The NES is an annual sample of employees whose NINO ends in 14, roughly 1% of employees, and contains payroll information for one particular week in April.[41] Data for individuals can be linked across years. Although only around 1% of individuals could be traced in the NES, the numbers entering WBTYP in any one year are large enough for even this small fraction to provide useful sample numbers.[42]

It would be possible to make some comparisons between TDS data on the employment rates and earnings data of MA leavers and LFS data for people who had not completed an apprenticeship. However these comparisons would be fairly crude because of differences in question wording, both for the outcome variables and for background variables such as ethnicity that might be used to refine the comparisons.

---

[40] In YCS Cohort 9, 73% of young people in AMA in the spring after the end of compulsory schooling said that they had got a place in education, work or training that they wanted, compared to 59% in other WBTYP.

[41] Employees earning below the tax threshold are not covered by the NES unless they work for a large employer.

[42] In 1999-2000 there were 258,400 starts in WBTYP in England and Wales, including 88,700 in AMA and 80,000 in FMA (DfEE 2000b, Table 1).

5.2 Surveys now in progress

5.2.1 YCS Cohort 10

YCS Cohort 10 reached the end of compulsory education in summer 1999. Sweep 1 was conducted in spring 2000 and Sweep 2 in autumn 2000; Sweep 3 is planned for spring 2002.

417 respondents to Sweep 1 of Cohort 10 were in AMA, compared to 544 respondents to Sweep 1 of Cohort 9. This is because the overall sample size is smaller, due to a particularly low response rate (55% compared to 65% at Sweep 1 of Cohort 9). The poor response is thought to be partly due to the extra length and complexity of the postal questionnaire, and is likely to increase the bias in YCS against less able young people. However, Sweep 2 of Cohort 10 had a better response rate than Sweep 2 of Cohort 9 (around 73% of Sweep 1 respondents compared to 66% for Cohort 9), mitigating some of the impact of the poor Sweep 1 response. This was partly because Sweep 2 was carried out only six months after Sweep 1 rather than after the usual interval of 12 months, and partly because of a greater use of telephone interviews. Sample attrition is likely to reduce the 417 Sweep 1 AMA trainees to around 200 by Sweep 3 and to perhaps 133 by a fourth sweep, compared to an estimated 170 at the equivalent stage of Cohort 9.

The number of young people in FMA at Sweep 1 of Cohort 10 is 189. Sample attrition is likely to reduce this number to around 80-85 by Sweep 3 and to perhaps 60 should a fourth sweep be held.[43] This is a very slender base on which to construct estimates of programme effects for FMA, but FMA trainees could be included in an estimate for MA as a whole. At a fourth sweep, combined sample numbers for those in either AMA or FMA at Sweep 1 are likely to be around 190-200, compared to an estimated 185 at Sweep 4 of Cohort 9.

Cohort 10 has a little more information on home background than Cohort 9. Additional variables include identifying young people living with step-parents and measures of the interest that parents took in their child's education. These may go some way towards reducing unobserved differences between MA trainees and comparison groups, but they are unlikely to make a large difference.

The possibility of using Cohort 10 to evaluate MA depends on whether a fourth sweep is carried out, as Sweep 3 is too early to collect information on outcomes for most MA trainees. All YCS cohorts have been surveyed at age 18/19, but only Cohorts 3

---

[43] This assumes a response rate of only 60% at Sweep 3, because the interval since Sweep 2 will be around 18 months rather than the usual 12, and a response rate of 73% at Sweep 4 should this be held six months after Sweep 3.

and 9 have had a sweep of the full cohort after this. The DfES allocates funding for YCS annually, and it is impossible to say at this stage whether there will be a Sweep 4 of Cohort 10. Should one be funded, it would not take place until autumn 2002.

Even if there is a fourth sweep of Cohort 10, the same problems would arise as with Cohort 9 over the exclusion from the evaluation of older entrants to MA.

5.2.2 Educational Maintenance Allowance Pilots data

The EMA pilot surveys are taking place in 15 Local Education Authorities (LEAs) where EMAs have been introduced and 11 comparison LEAs. The geographical distribution of these is somewhat arbitrary, with a heavy concentration in the North East, the North West and the Midlands.

Altogether four cohorts of young people sampled from Child Benefit records are being surveyed. Cohort 1 of the main EMA pilots is the only cohort for which a follow-up in the fourth year after the end of compulsory full-time education is currently planned - at the moment the last sweep of Cohorts 2-4 will be in the third post-compulsory year, which would be too soon to measure outcomes from MA. Cohort 1 reached the end of compulsory full-time education in summer 1999, so includes entrants to both AMA and FMA. Wave 1 took place during November 1999 - March 2000 and Waves 2-4 are planned to follow at yearly intervals, so data would not be available for the evaluation study until late spring 2003.

Certain problems involved in using Child Benefit Records as the sampling frame mean that there may be a bias in the original target sample against those who left full-time education early. For those who leave full-time education, details are collected of jobs and training. At Wave 1, information was collected by means of face-to-face interviews with young people and their parents. Waves 2-4 will use telephone interviews, supplemented by face-to-face interviews with those not reachable by telephone. This means that information that could be used to control for selection effects is richer and of higher quality than in YCS, which uses mainly postal questionnaires.

At Wave 1 useable data was obtained on 9,803 young people, representing a response rate of 72%. Waves 2-4 aim for a response rate of 80% of those in the eligible income group. At Wave 1 373 respondents were in AMA and 325 in FMA. Making the most optimistic assumptions about eligibility and response rates, this would yield Wave 4 information on 191 young people who were in AMA at Wave 1 and 166 young people who were in FMA at Wave 1. Design effects arising from the highly clustered nature of the sample will reduce the effective sample size further. However

if fourth sweeps of Cohorts 2-4 were funded, sample numbers could be boosted by aggregating across cohorts.

There is no way in which the EMA Pilots sample can be considered a nationally representative sample, and this creates difficulties when generalising from the survey results on MA. Further complications are introduced by the fact that selection into MA is likely to operate differently in the EMA pilot and comparison LEAs. However EMAs are likely to encourage some young people who would otherwise take work-based routes, including MA, to stay on in full-time education after 16. There is thus a possibility of using residence in an EMA pilot LEA rather than in a comparison LEA as an instrumental variable to control for selection effects between MA participants and young people in full-time education (see Section 3.3).[44]

Like YCS, the EMA Pilots data offers the possibility of defining several different comparison groups for MA. However it suffers from the same problem of excluding older entrants to MA.

Cohort 1 of the EMA Pilots are the same age as Cohort 10 of YCS, the sweeps take place at approximately the same intervals, and there are similarities in the form of words used to ask about WBTYP. Although it would not be possible to aggregate data from the two studies because of the differences in their design, parallel analyses could be carried out on them in order to see whether the two data sets supported the same inferences.

5.3 Possible future data sources

5.3.1 Longitudinal Survey of Young People

The DfES proposes to launch a new longitudinal study of young people, the design of which is still under discussion. The design recommended by the original feasibility study involves sampling young people from school rolls in Year 9 (age 13/14), with subsequent sweeps in each subsequent year apart from Year 10 up to age 20/21, with a possible three further waves following the cohort up to age 23/24.[45] In Year 9, face-to-face interviews would be conducted with both the young people and their parents, and data would also be collected from schools. There would be further face-to-face interviews with the young people late in Year 11, just before they reached the end of compulsory education, with more data added from school records. Subsequent interviews would be by telephone. This design may be modified in order to obtain adequate sample sizes for young people in minority ethnic groups.

---

[44] We are grateful to a participant in the April 5 2001 workshop for this suggestion.
[45] La Valle and Shepherd 2000.

The initial sample size recommended by the feasibility study is 21,000, which is predicted to yield a sample of 11,000 at age 21/22. This estimate relies on achieving much better response rates than in YCS, but this may be possible if face-to-face and telephone interviews are used rather than postal questionnaires, and more effort is put into keeping contact with the sample and tracing movers. If this level of response is achieved, then sample numbers for both AMA and FMA will be ample for evaluation purposes, particularly as a follow-up at age 21/22 would allow 18-year old entrants to AMA to be included in the evaluation.

One of the main purposes of the study would be to track young people through education, training and work after age 16, so it is likely to collect all the information an MA evaluation study would need on post-16 routes. It is also intended to collect detailed information on educational aspirations and school experiences in the early teenage years, on family background and on a range of measures of social disadvantage. It should also be possible to link in information held on the IPR in order to get results in the school Standard Attainment Tests (SATs) and in GCSEs, A and AS levels and GNVQs. Thus the study will have a much stronger array of variables to control for selection effects than any of the data sets considered so far, with the particular advantage that they will all be collected before the time of entry to MA.

This survey would be very suitable for measuring the impact of MA on the job chances and earnings of trainees, as it would permit MA trainees to be matched with young people taking other post-16 routes on the basis of a wide range of characteristics measured in Year 9 or Year 11. However the final design is not yet settled and the first sweep will not take place until autumn 2002 at the earliest. It is possible that two cohorts would be sampled at that time, one in Year 9 and one in Year 11. Thus no data would be available for the evaluation study until spring 2007, when the Year 11 cohort would be aged 19/20, the earliest age at which outcomes could be measured. Though this means a long wait for results, it also opens up the possibility of ensuring that information is collected on all the variables needed for the evaluation. The extent to which older entrants to MA could be included in the evaluation would depend on there being continuing follow-up surveys into the mid-twenties, which of course involves an even longer wait for results.

5.3.2 Modifications to the LFS questionnaire

It might be possible to expand the LFS questions on apprenticeships to identify those who had started an apprenticeship but left without completing, and to distinguish AMA from FMA. However, both these would require additional questions to be inserted in the questionnaire, and given the competition for space it is unlikely that

they would be approved. The value of such a modification is in any case limited, given the other problems in using the LFS for evaluating MA (see Section 5.1.2). If new questions were included, the earliest quarter in which they could be included is probably spring 2002.

5.3.3 Additional sweeps of YCS

Information on longer-term outcomes from MA could be obtained from YCS if there were a sweep in the early twenties. This would also permit young people who enter MA at older ages to be included in the evaluation, thus increasing potential sample numbers for MA and reducing the limitations on coverage. However Cohort 9 will not reach age 21/22 until 2002/03, and Cohort 10 will not reach this age until 2004/05. Furthermore, sample attrition is likely to be a serious problem. Response at the 23/24 year old sweep of YCS Cohort 3 represented 62% of response at age 18/19. Since then, YCS response rates have generally declined substantially, and we could not expect a late sweep of Cohort 9 or 10 to do as well.

5.3.4 An entry cohort from the Trainee Database System

The discussion so far suggests that the proposed new longitudinal study discussed in Section 5.3.1 offers the only real possibility of estimating the net effects of MA on individuals' job chances and earnings with any degree of confidence. Unfortunately the timeframe for this option is extremely long, and falls outside the most distant of the four dates by which the DfES needs to have results. The DfES has indicated that if it is not possible to measure net MA effects within the specified timeframes, then the feasibility study should consider a study with more limited aims that would nevertheless increase the information that we have on outcomes from MA.

A new survey of a sample of entrants to WBTYP drawn from the TDS might serve this purpose. This could enhance the data already available in the TDS in a number of ways, and by including entrants of all ages, would avoid one of the main limitations of using an age cohort for the evaluation. Although such a survey would not allow us to estimate the effects of MA relative to young people outside of WBTYP, it would permit comparisons between AMA, FMA and OT.

The policy interest of comparisons between MA and OT is limited by the fact that OT is due to be phased out. Nevertheless the comparison could indicate whether the better-structured and higher quality training offered by MA produced better outcomes than OT.

There also appears to be limited policy interest in a direct comparison between AMA and FMA. Where these are offered in the same sector by the same provider the

differences between them may not be very great, and may be more a question of whether the trainee is considered capable of Level 3 or only Level 2 qualifications than of differences in the actual training provided. As the MA programme develops, it is planned that AMA and FMA should become increasingly inter-linked, with easier progression between them. However analysis of the new survey could focus more on outcomes from the achievement of qualifications of different levels, which research suggests have different values in the labour market, than on the formal distinction between AMA and FMA.[46]

A new survey would also make it possible to compare outcomes for MA participants who complete their training and those who drop out early. There is currently very little information about this, though the very high wastage rates from WBTYP (see Section 4.6) make it an important issue.

It is hard to see how the design of such a new survey could include a new survey of a comparison sample of non-participants in WBTYP. However it would be possible to make some comparisons with groups identified in the LFS. This could be achieved by ensuring that the new survey collected information in a way that was consistent with LFS definitions, and included all the standard LFS questions that are used to model earnings. Thus outcomes for MA trainees in terms of employment rates and earnings could be compared with people identified in the LFS who had not completed an apprenticeship, or who had completed an apprenticeship outside the MA initiative, after applying all the potential control variables available in the LFS. These include age, sex, ethnic group, marital status, GCSE results, health and disability, and region of residence (see Section 5.1.2). Although these control variables are too limited to mop up all relevant differences between MA trainees and non-participants, the comparisons would significantly enhance the information currently available about earnings and employment outcomes from MA. The analysis could be further refined (subject to some reasonably plausible assumptions) to compare outcomes for people with vocational qualifications of the same level acquired through MA and by other routes.

The viability of a new survey of a cohort of MA entrants is explored in more detail in the following section, which outlines a possible survey design.

---

[46] We are grateful to participants in the workshop on April 5 2001 for these points.

6 A NEW SURVEY OF MA ENTRANTS

6.1 Objectives

As noted in Section 5, there are limitations to all of the available existing data sources which could potentially help to evaluate MA. As a result, a new data collection survey will be required for a robust evaluation. In particular, a new survey will offer the following enhancements to the data already available in TDS:

- by covering a cohort of entrants based on start date, it would provide consistent analysis of outcomes;
- it would provide more detailed and better quality information on outcomes, including for example hourly earnings;
- more and better information could be obtained, to control for selection effects;
- in obtaining a higher response rate than the postal survey of leavers, it would reduce the response bias in existing information on outcomes.

Section 3 has noted the difficulties of possible research designs, with a matched comparison group design potentially the most viable. However, as discussed in Section 4, there are significant barriers to identifying an appropriate comparison group. Realistically, a suitable, single comparison group could only be included by one of the following methods:

- using a prospective design, and tracking a cohort of young people through time;
- utilising the forthcoming longitudinal survey of young people;
- screening the general population in order to obtain a sample of people based on specific characteristics.

The first two approaches are very much long-term solutions; they would not provide analysis in time even for the September 2005 deadline, so we have not considered these options in any detail at this stage. We would expect the third option to be prohibitively expensive, given the nature of the screening exercise, and the need to obtain a large sample, including separate sub-samples of different groups.

As a result, we would recommend a design which includes:

- Separate analysis of AMA and FMA participants. Although reforms to MA are leading to less clear distinctions between the two, as things stand, we would recommend a design which permits separate analysis. In particular, previous

experience has shown significant variation in outcomes according to attainment of NVQ level 2 as compared with NVQ level 3 qualifications.

- A sample of OT participants to be included in the new survey, for comparison purposes. Although the number of entrants to OT is declining rapidly, for existing cohorts this does permit meaningful comparison with the MA route. The new sample would therefore comprise entrants into three strands: AMA, FMA and OT.
- The facility to match MA participants to groups defined in the LFS, and to compare the estimates of job chances and earnings between these groups. These groups will need to be defined further, and this approach would only provide a partial comparison with MA participation, as the groups identified by the LFS would not cover all of the individual comparison groups (for example the NEET group).

This approach would identify differences in outcomes between three training groups (AMA, FMA and OT), as well as with the other groups defined in LFS; specifically whether differences were wholly due to the different characteristics of young people on the different schemes, or whether there were differences in added value in terms of earning potential. The limitation of this approach is that, while it permits analysis of the relative impact of different routes, it would not provide a comparison outside of these groups, and therefore permits only a partial estimate of the counterfactual case. However, as noted above, this is the most productive design available within the timescale.

6.2 Design

Following the overall design outlined above, we now need to consider whether a longitudinal study using cohort or panel data is needed, or whether a cross-sectional study with appropriate retrospective data would be appropriate. A key issue here is the timing of the survey interview, relative to participation in MA; specifically:

- In order to analyse the impact of MA on participants, a sufficient time needs to have elapsed for most if not all respondents to have completed their MA, and for meaningful outcomes to be measured.
- Against this, the greater the gap since leaving the course, the greater the problems of sample attrition in any survey.
- In addition, there have been developments in MA which impact on the relevance of earlier cohorts' experiences to current and future policy. In particular, FMA was only introduced in 1998, and the numbers on AMA, FMA and OT have changed over time, with numbers of starts on AMA levelling out since 1998, a

consistent increase in numbers entering FMA since its introduction, and a corresponding fall in numbers on OT.

Inevitably, the research design will need to balance these conflicting issues. In considering the design, we would note the following:

- We would intend to sample a cohort of participants based on entry date. The DfES's regular postal surveys of WBTYP participants have been triggered by leaving date, as that is the significant event for that exercise, but a consistent analysis of participants' progress will rely on a consistent start date. In this context, it is also worth noting that the details on leaving are missing or incomplete for many participants on the TDS (this may be because there is no link between the payment system and return of Section 3 of the Record Form). In the AMA extract we have analysed, 26% of April 1998 entrants do not have a completion date (a far higher proportion than may actually not have completed). This re-iterates the need to base analysis from the point of entry rather than leaving.
- For comparable analysis, we would need a consistent entry date between the three sample groups: AMA, FMA and OT participants; and for the timing of fieldwork also to be consistent.
- Entrants of all ages and from all sectors would be included in the sample.
- Interviews among a single cohort would need to take place at least three years after start date, in order to assess meaningful outcomes. However, as noted above, the recent introduction of FMA means that the 'start date' of any cohort could not be before 1998.
- A longitudinal survey among a cohort would have limited benefit and low value for money. Given that the first interview needs to take place at least three years after entry, there would be little to be gained from a second wave interview with the same population unless it occurred some time after that. However, a long gap between waves (translating to five or more years after entry) would be extremely difficult, given sample attrition
- The use of more than one cohort in the evaluation as a whole is worth considering, although clearly there are timetable implications. This would be a possibility for the September 2005 reporting deadline.
- The survey would need to compare participants of AMA, FMA and OT, so the sample needs to cover sufficiently large groups of all three.
- Any survey needs to obtain a reasonable response rate, which will be difficult, given sample attrition and the difficult nature of the population concerned (young and mobile). After allowing for sample attrition (discussed further in 6.5), a response rate of around 60% should be feasible.

6.3 Options for data collection methods

6.3.1 Face-to-face interviewing

We expect that face-to-face interviewing will be required for this survey. As well as offering the advantage of data quality, a face-to-face method is generally preferable for interviews that are longer than 20 minutes – we would expect this to be the case given the required coverage. In addition, face-to-face fieldwork is the most successful method for tackling sample attrition. However, a face-to-face approach is likely to be fairly expensive, particularly with a relatively large sample size. Costs would be contained by incorporating some clustering of addresses into the sampling process, although this depends on the number of cases which need to be sampled, the number available in the eligible sample frame, and the geographical distribution of these addresses. In general, fieldwork is more efficient with a clustered sample but a design effect is introduced. An unclustered design has an impact on fieldwork costs, as this affects the amount of time interviewers spend travelling between addresses, in turn restricting the number of interviews that can be achieved per fieldwork shift. The positive side of this design is that it is a 'high quality' sample design, without the design effect introduced by a clustered sample.

As far as response rates are concerned, we would generally recommend face-to-face interviews in order to maximise response rates. Further issues relating to response rates and sample attrition are covered below.

6.3.2 Telephone interviewing

Telephone interviewing would be the main alternative to a face to face method. Generally speaking, telephone interviewing is more cost effective and quicker to carry out. The coverage of telephone numbers on the database is also encouraging. The disadvantages are the likely interview length and the need to trace people who have moved – tracing respondents is certainly possible with telephone fieldwork, although it would be more successful with a face-to-face method. A further disadvantage is that despite the high proportion of telephone numbers in the database, a proportion will be missing or incorrect. It is possible to increase this proportion, by using a telephone look-up agency to add telephone numbers to names/addresses where they are not on the sample. For this respondent group, we would not expect a large proportion of missing numbers to be obtained; this is because the search is based on listed telephone lines. If the number is not listed, or if the respondent is not the named 'owner' of the telephone line, the match will not be successful. For this age group, many people will not be the named owners, as this will often be in a parent's name (if they live at home), or the record may not have been updated if the person has moved recently.

We consider using telephone contact (see Section 6.5), although not for the main interview.

6.3.3   Postal questionnaires

The main disadvantage of a postal survey is the likely response rate, and differential response between different groups. Given the experience of the TDS leavers surveys, we would not recommend a postal survey for evaluation purposes.

6.4 Sampling and sample size

The DfES's TDS would need to form the basis of any sample of MA participants. Previous experience and analysis of an extract of the AMA database suggest that there should be no serious problems in using the database as a sampling frame. The sampling procedure would have to take account of transfers between AMA, FMA and OT, to avoid double sampling.

The overall sample size is likely to be relatively large, given the desirability of substantial sub-samples of AMA, FMA and OT participants, as well as other sub-groups for analysis, such as different employment sectors and regions. It is likely that an achieved sample size of 4,000-5,000 will be necessary, in order to analyse the three participant groups separately. Alternatively, the sample size could be reduced by excluding the OT participants and/or combining the participant groups in the analysis. An achieved sample size of around 2,000 would then be acceptable.

As discussed above, the survey will need to use a cohort of participants with a consistent start date. Following discussions and analysis of an extract of the AMA database, we would recommend an 'annual' cohort, for example from September 1998 to August 1999. This would ensure that the sample covers a full representation of participants in terms of circumstances and entry routes, and would level out any seasonal variations in entry. The one disadvantage is in the resulting timing: for example, with a cohort spanning September 1998 - August 1999 (the earliest that could be used given the timing of FMA's introduction), fieldwork would need to span autumn 2001 to summer 2002, if a three-year gap is to be maintained between entry and interview. A later cohort (for example September 1999 - August 2000) would mean that fieldwork would continue until summer 2003.

In order to maintain a consistent gap between entry and interview, the fieldwork would need to be staggered. This could take the form of rolling fieldwork (with each month's sample of the entry cohort being interviewed in a corresponding month), or

in a batch format (for example three batches of fieldwork, with September - December entrants interviewed in the first batch of fieldwork, and so on).

These two sample cohorts (September 1998 - August 1999 and September 1999 - August 2000) are the most likely possibilities for a new survey. The former would permit earlier analysis (this is discussed further in Section 6.9 below), but this cohort would be at the beginning of FMA's introduction, and may therefore not be representative of FMA participants.

6.5 Tracing addresses

Maximising the response rate is essential to obtain reliable data. Previous surveys, both with the MA group and other samples of young people, have had low response rates. The difficulty of obtaining up-to-date addresses is often one of the main reasons for this.

- We would expect a high level of sample mobility in this group, and this is reflected in the response rates to the leavers surveys, and the number of 'invalid' returns – typically around 10% of the total sample, including letters returned by the post office, 'not known at this address', etc. The survey as envisaged is likely to be using addresses which, in many cases, are more out-of-date than in the TDS's regular leavers' survey, so this problem will be increased. From our analysis of the AMA data extract, of those who have left/completed WBTYP, 30% left after less than six months, while 10% left after two years; the average length of stay is 69 weeks for AMA, 31 weeks for FMA and 49 weeks for OT. The range in length will be an issue for the analysis, and also for sample attrition; it means that the most recent address for many respondents will be several years out of date.

While recognising the problem, it is difficult to speculate on the number of respondents who would be living at the same address at the time of the survey. This could be critical to the success of the survey, since it would be acceptable if around 15% of the sample could not be traced, but if the proportion were considerably higher (for example, 25% or above), it could jeopardise the survey. As a result, there are several possibilities which are worth considering.

- A small-scale exercise to check on addresses and telephone numbers. A small sample of cases could be selected (for example, 100 from each of AMA, FMA and OT) from 1998/99 starts. These could be telephoned to check whether the telephone number and address were still correct. This would not provide precise figures, but would give a clear indication of the scale of the problem. It would be

most straightforward if this was carried out internally by DfES; the use of another organisation would create the need for an opt-out procedure because of data protection issues. If the number that had changed address was very high (for example, more than one in three), this may impact on the feasibility of the type of study that we are suggesting. If it is lower, it would provide useful information to feed into design and response rate assumptions. This would also give an indication of the likely success of a tracing exercise in the main survey; for example, if the telephone number or address were that of the parental home, it is likely that a new address or telephone number could be obtained. In the main data collection survey, we would recommend the inclusion of both of the two addresses on the TDS (from the time of entry and of leaving); given that the 'entry' address may be the parental address, this may actually be more productive in tracing the respondent than the 'leaving' address, for people who have subsequently moved.

- Telephone stage fieldwork in advance of the main data collection stage. In the event of a face-to-face survey three years after entry, one possibility would be to carry out a preliminary telephone interview. This would only be of use if this could precede the main survey by at least six months. The purpose of this exercise would be to check whether the respondent was living at the address, and to attempt to obtain new details if they had moved. Where the number/address was correct, the interview would establish the likelihood of moving in the near future and obtain an alternative contact (friend or relative). This interview could also include some basic details (for example work status), although it should be kept as short as possible.

- In the sampling stage of the main survey, cross-referencing between the AMA, FMA and OT databases (again, using NINOs), as leavers from one may appear in the database for another with a more recent address.

- Updating the database addresses from alternative sources. The sample could be cross-referenced with other government databases, for example the New Deal evaluation database, JUVOS, JSAPS or LMS, in order to try to obtain a more up-to-date address. As all these databases include NINOs, such cross-referencing could be carried out with confidence. Potentially, this approach could reduce the attrition rates in the sample. In particular, the cases that are likely to be updated in this way are also likely to be early leavers and more mobile respondents, where attrition rates will be highest. If possible, it would be useful to trial this approach, using a sample from the TDS.

6.6 Other ways of reducing sample attrition and increasing response

In addition to maximising the number of correct addresses traced, it will be important to put some effort into reducing attrition and maximising response rates, as the respondent group is young, mobile and potentially unpredictable. However, the survey will be highly relevant to them, and therefore could potentially obtain a reasonable response rate. The main stage survey will need to use a range of methods to achieve this.

- Attention to detail/guidance in:
  - Interviewer briefings,
  - Wording in advance/opt-out letters,
  - Scheduling fieldwork, including efficient allocation of cases, flexibility in using different interviewers to follow up on each respondent,
  - Identifying effective fieldwork during piloting.

- Interviewers' efforts in the field, by approaching the new occupant at the address or neighbours.

- Use of a freephone number for respondents to ring in with new contact details. This can be useful in cases where people at the old address are reluctant to pass on details, but are in touch with the mover and willing to pass on the request for them to get in touch with new details.

- If a preliminary telephone survey is carried out, there are various means of attempting to contain sample attrition between waves, through respondent letters, telephone calls and respondent newsletters. Change of address cards are often used on similar surveys to allow respondents/recipients to inform the survey organisation of their new contact details.

- Incentives are not traditionally used on this type of survey, although this is an issue which may need to be considered. However, the main problem for non-response is likely to be sample attrition, and incentives will not alleviate this problem. On balance, therefore, this is likely to be an expensive option for relatively little 'return'.

- Use of proxy interviews for repeatedly unavailable individuals. Many questions on a new survey are likely to be factual, concerning basic details relating to very 'visible' (to other people) aspects of life, e.g. whether working. People living in the same household as the individual (or even in a different house if close relatives) are likely to know the answers to many of the questions. We could

therefore consider the use of proxy interviews with people in certain specified relationships with the named person. Obviously a threshold would have to be reached before this option was allowed, such as a specified number of broken appointments or unsuccessful attempts to make contact with the individual. This option should be considered seriously, so that at least partial information can be obtained on participants. This approach is successfully used on other government surveys.

6.7 Target response rates

At this stage, we would suggest that the study should aim to keep the proportion of sample members who cannot be traced to no more than 20% of the original sample, and that the response rate from the remainder should be around 60%. This would yield an overall response rate of just under 50%. This would be a significant improvement on the response rates from the postal surveys of leavers, which range from 17% to 39%, typically around 25% for AMA, 20% for FMA and 32% for OT.

As well as the need to obtain a reasonable response rate overall, there is also the problem of differential response, which is inevitable to some extent, but will need to be contained as far as possible. One advantage for the survey is that other variables included in the TDS (age, sex, disability, ethnicity, previous WBTYP, and details of the planned training) will permit an analysis of differential response, so that appropriate weights can be applied. For a proportion of the sample, there will also be information from earlier leavers surveys, which could give some indication of how outcomes affect differential response.

6.8 Questionnaire coverage

A certain amount of information is already available in the TDS, and the questionnaire should not attempt to duplicate this, particularly for details about which the administrative record would be expected to be more accurate, such as start and leaving dates, SOC of training occupation and sector code. We would envisage the following topics being covered in the interview:

- experience of and views on WBTYP, including reasons for starting;
- activities before WBTYP, including work experience and training;
- whether completed the training agreed on entry;
- skills and qualifications gained;
- reason for leaving and immediate destination on leaving;

- earnings and hours of work in current job, and in most recent job for those not currently in work;
- details of current job, if any, including date started, occupation, industry, workplace size, training, whether belongs to a trade union and other factors that predict earnings;
- indicators of the likely sustainability of the current job, such as whether temporary or permanent, whether involves anti-social hours, learning opportunities and job satisfaction;
- details of any other jobs held since leaving training;
- job search activities of those not currently in work;
- work motivation;
- background factors correlated with outcomes, including qualifications before entering WBTYP, truancy at school, exclusion from school, literacy and numeracy problems, disability and health problems, household details and socio-economic status at age 15, including housing tenure and parental occupation and employment status.
- geographical information needed to merge in data on the local economy.

Where appropriate, the wording and coding of LFS questions would be followed, in order to facilitate comparisons with the LFS.

6.9 Costs and timing

The overall costs of a new survey on the lines described above are likely to be in the region of £250,000, based on an achieved sample size of 4,000-5,000. A reduced sample size of 2,000 (if separate analysis of AMA, FMA and OT participants is not required) would result in costs of around £150,000.

The September 2001 and January 2002 dates do not allow sufficient time for design, fieldwork and analysis of a new survey, but the September 2003 and September 2005 dates provide sufficient scope for a study. We have considered the timing implications of the two suggested cohorts.

Using a September 1998-August 1999 cohort, main fieldwork would take place between September 2001 and August 2002, which would result in analysis and reporting in time for the September 2003 date. Subsequent cohorts could follow a similar timeframe, for example a 2000/1 cohort interviewed in 2003/4. Using a September 1999-August 2000 cohort, fieldwork would take place in 2002/3, so reporting would not be in time for the September 2003 date, but could be carried out in 2004.

7 OTHER RESEARCH AIMS

7.1 Introduction

In addition to the net impact of MA on participants' long-term job chances and earnings, the DfES wishes to consider the feasibility of evaluating the long-term benefits of MA from the perspective of employers or the wider economy and of measuring variations in the effectiveness of MA between different sectors of the economy. This section of the feasibility study identifies under these broad headings the outcome measures that are likely to be affected by MA and the practicability of quantifying the net change in these outcomes due to MA. In addressing the feasibility of an evaluation from the perspective of employers, we generally view employers in aggregate as part of the wider economy. We do however briefly discuss the possibilities for evaluating the benefits of MA to individual employers. Throughout this section MA refers to both FMA and AMA.

7.2 Evaluating the net impact of MA from the perspective of employers

Employers play a central role in determining the success of MA. They are expected to take on modern apprentices, to provide on-the-job training and to contribute to the cost of off-the-job training. Also, other than the young people who undertake a modern apprenticeship, employers are likely to be the main benefactors of the policy. Hence, an assessment of the benefits of MA to employers is clearly warranted. Here we outline the main effects of MA on employers and the possibilities of quantifying these. It is helpful in this discussion to distinguish between individual employers and employers in aggregate as part of the wider economy.

7.2.1 The impact of MA on employers

In thinking about the possible effects of MA on employers, employers may in some sense be regarded in the same manner as the individuals who participate in MA. Employers incur a training cost, which is partially subsidised, in return for a more productive workforce and ultimately a gain in profitability and competitiveness. While the literature provides much evidence on the returns to training for individuals, in terms of earnings gains, it provides rather little evidence on the return to employers, in terms of productivity and in particular profitability gains. This is due in part to difficulties in obtaining data on firm productivity and profitability and the extent of training undertaken by the firm (Blundell *et al.*, 1999). Nevertheless, the fact that

employers do invest in staff training suggests that this activity is associated with a positive return despite the scarcity of hard evidence.[47]

The returns to the employer of hiring and training modern apprentices will generally depend on the cost of providing the training, the productivity gain associated with the training, the extent to which this productivity gain is specific to the employer, and the general process of wage determination. The productivity gain is in the first instance a mutual gain shared by the individual and the firm. However, if skills are easily transferable to another firm, the worker will be able to bargain for a greater share of the gain in terms of higher wages, leaving the firm with an improvement in productivity but little improvement in profitability. At the industry level the important factor is whether skills are industry specific.

The training provided by MA is to an extent job-specific, hence if individual productivity is raised by the MA training, it is likely that employers will capture and retain some of these gains in terms of improvements in profitability and competitiveness. This is supported by a survey of modern apprentices, which found that the majority of young people expect to stay with the same employer on completion of their apprenticeship.[48] Firms will also be affected by the impact of MA on the wider economy. For example, Haskel and Martin (1996) suggest that skill shortages at the aggregate level, rather than at the industry level, raise the skill premium and reduce the bargaining power of individual employers. If for example the provision of MA raises the supply of skills in the economy, this is likely to put downward pressure on the skill premium, raising the bargaining power of employers and allowing them a greater share of the productivity gain in terms of profits. These and other wider effects of MA of relevance to employers are discussed in section 7.3.

7.2.2 *Measuring the impact of MA on employers*

Here we discuss the possibility of quantifying the direct effect of MA on individual employers, leaving a discussion of the possibilities of quantifying the wider effects of the policy relevant to employers to section 7.3.2. As is clear from the outset, given the difficulties involved in obtaining data on firm productivity and profitability as well as the share of modern apprentices in total firm employment, the feasibility of conducting an evaluation on the firm specific benefits of MA is limited. These data difficulties are further compounded by the evaluation problem itself.

In principle, the ABI respondents database (ARD) and company accounts data could be matched to the information on training provider in the TDS, although there are

---

[47] Blundell *et al.* (1999), Skills Task Force (2000), and Barrett *et al.* (1998) provide a review of the evidence to date. See also Dearden *et al.* (2000).
[48] Coleman and Williams (1998)

likely to be some difficulties in doing so in practice.[49,50] First, the training provider on TDS is recorded by name and address, and should indicate the enterprise or organisation in which the MA trainee is employed (if indeed he is employed). However, it is not always clear that it does. Second, there are likely to be difficulties in the actual linking process, even if the training provider does indicate the employer correctly. If the training provider on TDS accurately records the company name and address, then in theory it should be possible to link the TDS to the Inter-Departmental Business Register, although permission to do so may be difficult to achieve. This would enable the TDS to be linked to the ARD providing establishment level data on value added, wages and salaries, employment and investment as well as the pay and number of MA trainees employed. However, the data would be limited to establishments in the production sector. Greater sectoral coverage could be achieved by linking the TDS to company accounts data by company name. Unfortunately, information on employment is somewhat sporadic in company accounts data, making it difficult to measure labour productivity and the proportion of MAs in employment.

The evaluation problem itself presents further complications. An evaluation of the effects of MA on individual employers would need to consider what employers would have done in absence of the MA policy. For example, would the employer have provided the same level of training anyway or would it have recruited individuals with vocational qualifications obtained through further education at zero cost? Having established this, an evaluation would need to find an appropriate comparison group of employers to identify the 'counterfactual' for employers training MAs. These are likely to be difficult to obtain given that for a number of occupations, MA is the predominant mode of entry (as discussed above in Section 4.2), and given the relatively high sectoral concentration of MA. For example, 32 per cent of young people in AMA in September 2000 were in Engineering Manufacturing, Business Administration and the Motor Industry alone, and 30 per cent of young people in FMA were in Business Administration and Retailing alone.[51] Variation across firms in the proportion of MAs hired, as obtained from the linking exercise described above, is likely to be of some assistance here. However, there is a limit to the information that can be gained, as it is likely that much of this variation is endogenous and can be explained away in terms of factors such as firm size, capital intensity and relative profitability.

Having highlighted the difficulties in obtaining data on individual firms and the severity of the selection problems in an evaluation of the impact of MA on individual employers, we would advise that it is neither possible within a reasonable budget, nor

---

[49] See Oulton (1997) or Griffith (1999) for a description of the ARD.
[50] Accounts data for 1500 companies quoted on the LSE or AIM are available on Data Stream. Accounts data for UK registered companies are available from BUREAU van DIJK electronic publishing limited (www.bvd.co.uk).

is it strictly necessary for the purposes in hand, as much of the information needed can be gleaned in the way suggested below. We suggest that an evaluation of the benefits of MA from the perspective of employers is conducted in the context of a wider economy evaluation. This is discussed in the next section.

7.3 Evaluating the net impact of MA on the wider economy

To the extent that MAs make a net difference to the productivity and employability of participants, it is likely that the policy will have wider economic implications and an evaluation of the benefits of the policy as a whole would require an assessment of these. An evaluation of the wider economy effects of MA may for example be used to inform an assessment of the long-term benefits of MA from the perspective of employers, as suggested above, and to feed into a cost-benefit analysis of the policy. An assessment of the wider implications is also useful in informing the evaluation of the policy's impact on participants' earnings and long-term job chances, as the wider implications of the policy in turn affect both participants and the comparison group.

MAs were originally introduced to improve the supply of intermediate level skills and form part of the current government's strategy to raise the share of the population with NVQ Levels 2 and 3 in meeting its 'National Learning Targets'.[52] Hence, if successful the policy should raise the quality of the national labour supply. While there are several aspects of MA that may determine its wider implications, such as its financing and occupational concentration, it is its effect on the quality of the labour supply that is key. There are two components to this. First, the effect of MAs on individuals. If individual productivity is not affected by participation in MA, compared to a counterfactual involving non-participation, then the wider economy effects of the policy will generally be absent. Second, the scale of the programme. Even if MA is very successful in raising earnings and the job chances of participants, it is unlikely to have significant implications for the rest of the economy if the number of participants is very small in relation to the aggregate labour force.

In the next section we set out how MA is likely to affect the wider economy based on our understanding of how the programme differs from previous policy and taking into account the magnitude of the programme in relation to the economy as a whole. The following section proposes means by which these wider economy effects may be quantified, discussing both the general methodologies available and their feasibility under different time-scales.

---

[51] Source: DfEE (2000b), Table 7.

[52] Department for Education and Employment (2000), "National Learning Targets Annual Report 1998/99", available from www.dfes.gov.uk

7.3.1 The wider economy effects of MA

The introduction of MA may be thought of as a change to the national system for the vocational education and training (VET) of young people. The national system for VET in place over the two decades before the introduction of MA is described by several authors (see for example Dolton, 1993, Layard *et al.*, 1994, and Prais, 1995). The general picture portrayed is one of a somewhat fragmented system that has led to an inadequate supply of labour with intermediate level vocational skills. The apprenticeship system in place at the beginning of the 1970s has suffered from the progressive abolition of the Industrial Training Boards (Layard *et al.*, 1994) and has never been developed to cover new and expanding sectors of the economy as has been the practice in continental Europe (Prais, 1995). The introduction of Youth Training Schemes (YTS) in 1983 and later Youth Training (YT) was intended to provide VET in a broader range of occupations, however, the skills obtained through these programmes were perceived by both young people and employers as being of a low quality and the main part of VET continued to be provided in further education (Layard *et al.*, 1994). A number of other training initiatives were introduced throughout the 1980s and early 1990s, however, 'the one common thread through all these initiatives has been the need to tackle youth unemployment' (Dolton, 1993, p.1265), rather than to tackle the skills gap.

Against this background, the introduction of MA and the revisions being made in light of the recent consultation process may be seen as an attempt at moving closer towards the continental system for VET.[53] It is intended to provide young people with an alternative to the academic route by establishing a unified system for VET, in occupations covering all sectors of the economy, through which they can progressively achieve nationally recognised qualifications of a high intermediate standard. Thus, if the policy is successful, it should raise the supply of high intermediate and intermediate level skills.

In thinking about the long-term effects of MA on the wider economy it is useful to think of the economy with MA in comparison to a counterfactual without. From the discussion above, the appropriate counterfactual is an economy generally lacking the VET infrastructure necessary to produce intermediate vocational skills. Below we discuss the likely implications of MA for growth and productivity, earnings, employment and competitiveness with this counterfactual in mind. If successful in raising the supply of skills, the benefits of the policy to the wider economy are important but are only likely to materialise in the very long term.

---

[53] http://www.dfes.gov.uk/ma.consultation

Productivity and growth

To the extent that MA participation enhances individual productivity, this should have implications for overall productivity. The effects on the wider economy will initially be small, but are likely to build up over time as the number of cohorts that have benefited from MA increases. There are generally two sorts of effects of MA on aggregate productivity - a direct effect via its effect on individual productivity and an indirect effect due to spillovers to training. An evaluation of the effects of MA on the wider economy would need to distinguish between these two effects. We discuss each of these in turn.

Starting with the assumption that there are no spillovers to training, the effect of MA on aggregate productivity basically depends on the share of individuals in the workforce who gain from MA and the size of the individual gain. By means of a very simple calculation we can gauge the magnitude of this effect on overall productivity and how it builds up over time. For example, if every year a proportion p of the workforce gains some qualification through MA that raises their productivity level by g per cent, then every year MA will raise aggregate productivity by around pg per cent. Thus, after the first cohort of MA graduates has been produced, aggregate productivity rises by pg per cent. The second cohort adds to this so that the aggregate productivity effect of MA rises to $2pg$ per cent, and so on. Eventually, the first MA graduates will exit the labour force, neutralising the impact on aggregate productivity of new graduates. Once this situation has been achieved, productivity will have been directly raised by Npg, where N is the length of working life.

In the last three years, roughly 88,000 young people have started an AMA in England and Wales each year.[54] Roughly half of these leave with a least an NVQ3 level qualification, representing around 0.12 per cent of the population of working age each year.[55] If this continued unchanged for the next 40 years, around five per cent of the labour force will have gained an NVQ3 level qualification through AMA in 40 years time. Thereafter, the proportion of qualified AMA graduates in the labour force would stabilise as the first graduates begin to retire. How much would this add to aggregate productivity? The answer to this depends on what young people and employers would have done had AMA not been available.

Evidence on the premium to individuals arising from participation in vocational training is mixed. Much of the evidence from the evaluations of youth training programmes suggests that the effects are negligible for the majority of youths (OECD,

---

[54] Source: Labour Market Trends, February 2001, Table F.2
[55] Source: Labour Market Trends, February 2001, Table F.5; In 1999-2000, 48 per cent of leavers from AMA left with at least an NVQ3 level qualification (survey respondents only). This is greater than in earlier years, however, as mentioned in section 2, the progress of earlier cohorts may be misleading.

1996). In an evaluation of the impact of participation in YTS, Dolton *et al.* (1994) find little evidence that participation raised earnings, except for the most disadvantaged. However, these studies do not necessarily provide the appropriate comparison to AMA. Advanced modern apprenticeships are intended to lead to high intermediate qualifications, and are intended to provide young people with a real alternative to the academic route. In this case a more appropriate indicator of *g* is likely to be the return to holding intermediate level vocational qualifications, as opposed to holding lower level vocational qualifications.[56] Blundell *et al.* (1996) estimate the increase in earnings as a result of holding intermediate compared to lower vocational qualifications at around five per cent. Evidence in Dearden *et al.* (2000) suggests that the productivity gain associated with vocational training may be twice as high as the earnings gain. These results would give a value of *g* (the direct productivity gain from participation in AMA) of around ten per cent.

The basic calculation above suggests that after forty years, AMA will have raised the level of aggregate productivity by at least half a per cent. After ten years, AMA will have raised aggregate productivity by an eighth of a per cent. On average, in terms of GDP at current levels, this corresponds to roughly half a billion pounds per year in the first ten years. These numbers would be even smaller if we took into account the fact that skills may depreciate over time (see Blundell *et al.* 1996). We outline a number of reasons to believe that the actual rise in productivity will be larger in the longer term than illustrated by this simple calculation, which an evaluation would need to take into account.

There are a number of reasons why the long-term productivity gains may be larger than suggested above. First, it is possible that participation will increase as MAs become more standard in sectors that do not have an apprenticeship history and if, unlike YTS and YT, MAs can gain the confidence of young people and employers. Second, it is possible that there are greater productivity gains if there are externalities to training. For example, the availability of skilled labour may raise the productivity of unskilled workers in the same workplace, raising the social return to training above the private return as suggested by Lucas (1988). This suggests that the aggregate productivity gain of MA may exceed the individual gain. Alternatively, a critical mass of skilled labour may facilitate learning-by-doing and technology adoption (as in Greenwood and Yorukoglu, 1997) or may aid process innovation, such that the employment of skilled labour directly affects the rate of productivity growth. Again this suggests that the aggregate productivity gain of MA may exceed the individual gain. If MA does facilitate learning-by-doing this opens the further possibility that MA may encourage R&D or physical capital investments. In this case the initial rise in productivity due to MA will be amplified by investment in other forms of capital.

---

[56] As mentioned earlier in this report, the appropriate comparison group is likely to be more complex.

Evidence of this effect of training on manufacturing investment is found in Nickell and Nicolitsas (2000).

In any evaluation of the productivity effects of MA on the wider economy, it will be important to consider such externalities and to demonstrate the sensitivity of the results to the particular assumptions made about the way in which these externalities are generated.[57]

Earnings, employment and competitiveness

If participation in MA involves a productivity gain to the individual, it is likely that this will be associated with an earnings gain. It may also be associated with a gain in competitiveness to the employer. The productivity gain referred to here is *net* of what the individual and the employer would have achieved in absence of MA, and hence net of any deadweight associated with the policy. Generally, the productivity gain will be split between the employer and the employee according to the nature of the bargaining process. Besides depending on the institutional set up, this will depend on the relative supply of skilled to unskilled labour (which in turn depends on relative earnings), aggregate unemployment, and the extent to which the skills obtained through training are general or specific (i.e. skill transferability). The greater the bargaining power of the worker, the greater the earnings gain and the smaller the reduction in unit labour costs and rise in competitiveness for the firm. In aggregate, a stronger bargaining position of workers will imply a greater rise in average earnings, and a smaller rise in employment and competitiveness.

As mentioned above, evidence on the earnings effects of training programmes for young people is mixed. However, the literature generally finds an earnings premium associated with possessing vocational qualifications, suggesting that workers do reap some of the gains from undertaking training. As discussed in section 7.2.1, workers are unlikely to be the only beneficiaries. Employers providing the training should also enjoy some increase in competitiveness. From an aggregate perspective there are other reasons to believe that a rise in the aggregate supply of skilled labour will manifest itself in employment and competitiveness gains as well as earnings gains. Several studies have found that in the last two decades skills mismatch has lead to lower levels of equilibrium employment (Manacorda and Petrongolo, 1999, Nickell and Bell, 1995 and 1996). A greater supply of skills would help to alleviate this mismatch, raising the equilibrium rate of employment and reducing unit costs. For example, in an estimated model of the UK economy, Riley and Young (1999) find that an increase in the supply of labour with higher intermediate skills, similar to the change brought about after ten years of AMA graduates in the simple calculation

---

[57] Particularly in light of the continuing debate about the importance of externalities to training.

above, would raise sustainable employment, and reduce the skill premium and unit labour costs.

However, the productivity gain associated with MA is not likely to improve aggregate competitiveness by much in the shorter term, irrespective of the bargaining position of firms. It was suggested above that it could take ten years to raise aggregate productivity by an eighth of a per cent, although it would not be entirely correct to use this number to infer the rise in international competitiveness. Initially, MAs were largely concentrated in the more traditional apprenticeship sectors such as manufacturing. Hence, initially, the productivity gain of the wider economy is likely to be smaller than the productivity gain of the manufacturing sector alone. The manufacturing sector represents the largest share of the traded sector, and hence the concentration of the productivity gain here will lead to a greater rise in international competitiveness than suggested by the rise in aggregate productivity. Research in Oulton (1996) suggests that firms would use this rise in competitiveness to raise market share rather than to boost profit margins, which would raise employment. These effects should also be taken into account in an evaluation of the wider economy effects of MA.

7.3.2 *Measuring the wider economy effects of MA*

The importance of taking into account wider economy effects in evaluating large-scale policy intervention cannot be underestimated. This is vividly illustrated by Heckman *et al* (1998), who show that the estimated impact of a tuition subsidy on college enrolment may be up to ten times smaller than standard studies suggest, once the wider economic effects of the policy are taken into account. Another illustration is provided by a recent study of the macro-economic impact of the New Deal for Young People (NDYP) (Riley and Young, 2000), which found that the direct impact of NDYP on employment and unemployment is amplified by around 30 per cent taking into account the impact of the policy on aggregate wage pressure.

It is clear from the previous section that the introduction of MA is likely to have wider effects on the economy if it is successful in raising participants' productivity and participation, however the effects are mostly long-term. Despite the importance of taking into account wider economy effects in evaluating large scale policy interventions, this is something that is rarely done (Heckman *et al*., 1999). The literature available on evaluating wider economy effects is thus scarce. Nevertheless, there are some general guidelines one may follow. One can attempt to quantify the effects of policy either by econometric investigation or by calibrated model simulation. For the purposes of evaluating the wider economy effects of MA, relying on econometric investigation alone is unlikely to be a fruitful evaluation strategy. In terms of the aggregate economy, the effects of MA particularly in the early years will

generally be too small to detect in this manner. Instead an evaluation will have to be conducted *ex ante*, in the form of an appraisal, relying on model based evaluations and simulations informed by empirical investigation. In the following we set out the main issues involved in an evaluation of MA of this kind. First we discuss key features of a model that could be used to evaluate the effects of MA on the wider economy. Next we discuss the possibilities for obtaining estimates of the direct effect of participation in MA on individual productivity.

Model features

The outcome of an *ex ante* evaluation of the wider economic effects of MA will be very sensitive to the way in which the evaluator perceives the relation between the policy and the macro economy. We have set out our view of the outcomes and relationships that need to be considered in an evaluation of the wider economy effects of MA. In particular, the relationships that need to be considered are the factors influencing the participation decision of both individuals and employers, the skill-specific wage determination process, the extent to which there may be spillovers from MA training and the sectoral distribution of productivity changes in terms of openness to trade. Any model of the wider economy constructed to simulate the effects of MA would thus need to be specific about these relationships.

The participation decision for individuals must depend on the expected return to participation compared to other available alternatives.[58] Some consideration should also be paid to the firm's decision to hire modern apprentices and provide on-the-job training. This decision should depend on the expected return to this activity and needs to be broadly consistent with the legislative framework in place. The process by which wages are determined is crucial for the effect of MA on employers versus individuals. Skill-specific wage determination could for example be based on bargaining rules standard in much of the literature (see for example Layard *et al.*, 1991). The bargaining power of employers versus employees should at least depend on skill shortages and aggregate unemployment. The evaluation may also take into account the extent to which the skills obtained through training are general or specific. A number of alternatives should be presented in order that the range of potential outcomes can be discussed.

Particular attention will need to be paid to the way in which human capital is built into the production function and the way in which human capital is accumulated, as the evaluation of the productivity effects of the policy will depend on these features. The evaluation will need to consider both the direct effect of the policy on productivity as well as potential externalities as discussed in section 7.3.1. For example, learning-by-

---

[58] In particular the alternatives that identify the counterfactual scenario for individuals.

doing effects or the innovation process may be modelled by including skilled labour as a determinant of total factor productivity growth (see e.g. Greenwood and Yorukoglu, 1997). Alternatively, human capital may be included as an ordinary factor of production. Aghion and Howitt (1998) provide an extensive review of the ways in which such effects may be modelled. Once again, a range of potential effects should be evaluated.

Besides these particular features, the model clearly needs to be built around a sensible long-run equilibrium framework, as the introduction of MA is a policy intervention intended to benefit the economy in the long-term. It would be advisable if the structure reflected the nature of production and competition in the UK, and specifically took account of the implications of imperfect competition for trade and output. Trade effects might mitigate the output expanding effects of MA, and these cannot be ignored. Hence, any model structure used should also have a full description of the trade and competitiveness relationships as well as the production function. It will also need to demonstrate an adequate description of the main trends in the UK economy given sensible parameter values informed by the literature. If detail on sectoral breakdown and the public finances is desirable, for example to evaluate the impact on international competitiveness and to feed into a standard cost-benefit analysis, the basic model will also need to be embedded in a more general framework.[59]

A large part of a potential *ex ante* evaluation exercise will be the investigation of the stylised labour market and the construction and use of a model with these main features. Resources would be saved if this evaluation would take place in the context of a pre-existing theoretically coherent model. Further resources would need to be set aside for the demonstration of the sensitivity of the evaluation results to alternative specifications of the key relationships outlined above and to alternative parameter values.

Estimates of the direct effect of MA participation on individual productivity

To quantify the impact of MA on the wider economy, an estimate of the individual productivity gain associated with MA participation is required to feed into the model framework discussed above. This could be obtained from an evaluation of the impact of participation on individuals' earnings. While this is an imperfect measure of the productivity gain, it is unlikely that one would be able to obtain direct estimates of the impact on firm level productivity (as discussed in section 7.2.2). To use the results from a study of individuals, the estimation in that part of the evaluation would need to

---

[59] An evaluation of the public finance effects of MA will require additional information on the exchequer costs of MA provision.

control for changes in relative skill supply, as well as a measure of the state of the economy, such as the ratio of vacancies to unemployment. This is to ensure the study provides estimates of the direct effect of participation. These controls would need to be at some geographical level of disaggregation, given the lack of the time dimension in this part of the evaluation. Care would also need to be taken to ensure that the counterfactual, upon which the estimate of the impact of MA on individual productivity is based, is used consistently in the evaluation of the wider economy effects. The individual level study would also be able to provide separate estimates of the impact in the traded and non-traded sector of the economy, for example manufacturing versus non-manufacturing. Given the differential developments in the demand for NVQ2 and NVQ3 level skills over the last two decades (see for example Machin and Van Reenen, 1998, and Haskel and Heden, 1999), it would also be useful if it could provide separate estimates of the impact of FMA and AMA on individuals.

An alternative source of information on the individual productivity gain associated with MA participation is the literature on previous training initiatives and the returns to vocational qualifications in the UK (e.g. Blundell *et al.*, 1996). It is also possible that evidence from training initiatives in other countries could be used. For example, Germany and France have long established apprenticeship schemes in most sectors of the economy (Prais, 1995). Information of this kind does not provide as adequate a base for an evaluation of the wider effects of MA as does direct evaluation evidence on the individual effects of MA. Nevertheless, if carefully conducted, a study based on this secondary information would provide a useful appraisal of the potential wider economy effects of MA even in absence of a study of the effects on individuals.

7.4 Variations in the effectiveness of MA between different sectors

The DfES has also expressed interest in measuring differences between sectors in the effectiveness of MA. In particular they have expressed an interest in measuring the variation in effectiveness between industrial sectors such as services and manufacturing and between traditional and new sectors of the economy. If differences in 'effectiveness' are interpreted as differences in the direct effect of MA on individual productivity, it is clear from the discussion in section 7.3 that these sectoral distinctions are likely to be useful in evaluating the impact of MA on the wider economy. First, the distinction between manufacturing and services is likely to highlight differences between the traded and non-traded sectors of the economy, facilitating an analysis of the effects of the programme on international competitiveness. Also, the distinction between new and old industries is helpful in determining the impact of MA on aggregate productivity. For example, learning-by-doing and technology adoption are likely to be important factors in determining the productivity effect of MA in new industries, but perhaps less so in traditional

industries. Sectoral differences in the individual productivity effect of MA are also important in determining the response of the skill premium to the extent that the sectors differ in skill intensity (Haskel and Slaughter, 1998).

7.4.1   *Measuring variations in the effectiveness of MA between different sectors*

It is in principle relatively easy to measure differences in the effectiveness of MA across sectors within the framework discussed in previous sections for evaluating the benefits of MA to individuals. The data sets recommended there contain details on industrial and occupational coding. Differences in the effect of MA across sectors could for example be measured by including interactive treatment and industry dummy variables in estimation. However, it would not be feasible to estimate sectoral differences at a more disaggregate level, due to the occupational and industrial concentration of MA trainees mentioned in section 7.2.2. Any evaluation of the wider economy effects should be able to allow for a broad disaggregation of the labour market by sector.

7.5   Costs and timing

The DfES wish to consider the feasibility of evaluating the long-term benefits of MA from the perspective of employers or the wider economy and of measuring variations in the effectiveness of MA between different sectors of the economy. The conclusions from sections 7.2 - 7.4 as to the feasibility of each of these are that

- an evaluation of the long-term benefits of MA from the perspective of employers is best achieved in the context of an evaluation of the long-term benefits of MA to the wider economy;
- an evaluation of the wider economy effects of MA would of necessity be conducted *ex-ante* by means of model simulation, and may either draw upon the results of the evaluation of the impact of MA on individuals' earnings and employment or on the literature on the returns to training;
- an evaluation of the variations in the effectiveness of MA between sectors can be conducted for relatively broad sectoral disaggregations and could be facilitated in the evaluation of the impact of MA on individuals at very little additional cost.

An evaluation of MA, providing estimates of its long-term impact on productivity and growth, earnings, employment and competitiveness, conducted along the lines sketched out in section 7.3, contains three main elements. These are model construction, obtaining estimates of direct effects on individual productivity, simulation and sensitivity analysis of the results to key parameters and relationships. We suggest that the DfES would need to budget around £80,000 for this work, which

could be conducted over the space of a year. This is purely a rough estimate, as the actual cost would depend on the details required of the modelling. For example, the DfES may wish to allow a further £20,000 for incorporating sectoral disaggregation into the modelling of the wider economy.

Clearly, the introduction of MA is a policy intervention intended to benefit the economy in the long-term. This does not however imply that an evaluation need be postponed for ten years. An evaluation of the longer-term effects on the economy would of necessity be carried out *ex ante* based on simulation. The timing of such a study is primarily dictated by the timetable for the evaluation of the impact of MA on individual outcomes, unless the study solely makes use of evidence from previous training programmes and the experience of other countries.

8 TIMEFRAMES AND OPTIONS

8.1 Introduction

The DfES requires recommendations for the best options to produce results within each of four separate timeframes which are chosen to fit in with the information needs of the biennial spending review. Here we summarise the options available within these timeframes for estimating the net impact of MA on the earnings and employment chances of individuals, and for estimating the wider economy effects of MA. For a fuller discussion of each option, the reader is referred to the appropriate section of the main body of the report.

8.2 Estimating the impact of MA on individuals

8.2.1 Results by September 2001

To get results within this very short timeframe we would have to use existing data sets. Findings would have to be treated as preliminary as there would be little time to check results. They would rely heavily on descriptive statistics, as there would be not much time to develop well-specified statistical models.

Preliminary analysis of the LFS (see Section 5.1.2)

This option would compare the employment rates and earnings of people who had completed a recognised trade apprenticeship under the MA programme with those of people who had not completed trade apprenticeships.

Advantages:
Cheap; very large nationally representative sample with big enough sample numbers for MA; very good data on current occupation and earnings; covers everyone who has completed MA, regardless of age.

Disadvantages:
No data on training sector, type of provision or date of completion; no information on trainees who left before completion; AMA only covered with a probable bias towards the traditional craft sectors; inadequate set of background variables to control for selection effects; identification of comparison groups limited by lack of information on education, work and training histories.

<u>Preliminary analysis of the TDS</u> (see Section 5.1.3)

This option would compare the earnings and employment rates of people who entered MA and people who entered OT at a point roughly six months after leaving training.

Advantages:
Cheap; covers AMA, FMA and OT and entrants of all ages and from all sectors; good data on training plans at entry; completers distinguished from non-completers.

Disadvantages:
Administrative data at the point of leaving sometimes missing; fuller data on outcomes based on a survey with a very poor response rate; all information on outcomes relates to a variable interval after entry; limited set of background variables to control for selection effects; no information on young people outside of WBTYP.

8.2.2 Results by January 2002

<u>Fuller analysis of the LFS and TDS</u>

With four extra months to complete the work, more careful analysis of the LFS and the TDS could be carried out and statistical models developed, with adjustments in the case of the TDS for response bias in the postal survey of leavers. In addition, it may be possible by this date (subject to data protection requirements) to link the TDS with JUVOS to get information on subsequent spells of claimant unemployment, and with the NES, to improve data on earnings. Otherwise the main advantages and disadvantages of the two data sources remain the same.

<u>Analysis of YCS Cohort 9 Sweep 4</u> (see Section 5.1.1)

This option permits the employment rates at age 19/20 of 16/17 year old entrants to AMA to be compared with those of young people who took alternative routes.

Advantages: Cheap; nationally representative sample; longitudinal data; good information on training and qualifications; reasonable set of background variables to control for selection effects; possible to identify several alternative comparison groups.

Disadvantages: Sample numbers for trainees too small to detect small programme effects or to measure differential effects for different sectors, types of trainees or types of training provision; no data on earnings outcomes; older entrants to MA not covered; analysis restricted to AMA as sample numbers for FMA very small; tight timetable given that data are not expected to be available until October 2001.

8.2.3 Results by September 2003

Analysis of YCS Cohort 10 (see Section 5.2.1)

This option depends on there being a sweep of the cohort at age 19/20. It would permit analyses similar to those that could be based on YCS Cohort 9, with the potential addition of information on earnings outcomes.

Advantages:
As for YCS Cohort 9, plus coverage of FMA; slightly more background variables than for Cohort 9.

Disadvantages:
As for YCS Cohort 9, plus no guarantee that there will be a fourth sweep; tight timetable to meet the September 2003 deadline dependent on data from a fourth sweep being made available by early summer 2003.

Analysis of the EMA Pilots data for Cohort 1 (see Section 5.2.2)

Like the analysis of YCS Cohort 10, this option would compare the employment rates and earnings at age 19/20 of 16/17 year old entrants to MA with those of young people who took alternative routes after the end of compulsory education.

Advantages:
Cheap; coverage of FMA; longitudinal data with good information on training and qualifications; richer and better quality data than in YCS; possible to define several alternative comparison groups; fourth sweep already funded; good data on earnings.

Disadvantages:
Small sample numbers for MA unless additional sweeps of the other EMA cohorts are funded; older entrants to MA not covered; highly clustered sample in arbitrarily chosen areas reduces effective sample size; generalisability of results limited by complicating factor of the EMA experiment; tight timetable to meet the deadline dependent on data from the fourth wave being made available by early summer 2003.

Joint analysis of YCS Cohort 10 and the EMA Pilots data

While it would not be possible to aggregate data from these two studies, the similarities in their timing and design make it possible to test whether the results in the one are replicated in the other, thus compensating somewhat for the small sample numbers for MA in each separately. The disadvantages remain as for the surveys separately.

Analysis of the LFS with expanded questions on apprenticeship (see Section 5.3.2)

This option would enable analyses based on the LFS to distinguish AMA from FMA and completers from early leavers. However it could be very difficult to get agreement on the new questions and there would be a very tight timetable to meet the September 2003 deadline. The other disadvantages of using the LFS remain the same as already outlined in Section 8.2.1.

Analysis of a 21/22 year old sweep of YCS Cohort 9 (See Section 5.3.3)

This option would permit an analysis similar to the analysis that could be based on Sweep 4 of YCS Cohort 9 (see Section 8.2.2), but would have information on longer-term outcomes and could include information on earnings. Entrants to MA up to age 18 could be included. However high sample attrition is likely to make this option unviable.

New survey of a cohort of WBTYP entrants (see Section 5.3.4)

This option permits the earnings and employment rates of AMA, FMA and OT trainees at a point three years after entering training to be compared with each other. The progress of completers and non-completers could be compared. Comparisons could also be made with the earnings and employment rates of LFS respondents who had not completed apprenticeships, or had completed apprenticeships that were not part of the MA initiative.

Advantages:
Based on a cohort of entrants rather than leavers so permits a consistent analysis of outcomes; large sample numbers; covers entrants of all ages; distinguishes early leavers; better information on outcomes and background variables than the TDS; better response rate than the postal survey of leavers; possible to add TDS data to get good information on training plans, sector and employed status; makes it possible to identify comparison groups in the LFS.

Disadvantages:
Expensive; likely to be problems with tracing the sample; timing permits only a three-year gap between entry and measurement of outcomes and the timetable is still tight for the September 2003 deadline; relies on retrospective data for background variables; control variables for comparisons with groups identified in the LFS limited to existing LFS variables..

8.2.4 Results by September 2005

New survey of a cohort of WBTYP entrants

The main advantages and disadvantages of this option remain as described above, but the longer timeframe permits a generous timetable and a four year gap between entry to MA and the measurement of outcomes.

8.2.5 Results by September 2007

Analysis of the new longitudinal study of young people (see Section 5.3.1)

This option permits the earnings and employment rates at age 19/20 of 16/17 entrants to MA to be compared with those of young people taking alternative routes.

Advantages:
Good sample size; rich and high quality data with longitudinal structure and measures of key background variables before age 16 make it very suitable for the matched comparison group methodology; possible to define several different comparison groups; covers both AMA and FMA.

Disadvantages:
Design uncertain; funding of subsequent sweeps in early twenties unlikely to be guaranteed; older entrants to MA not covered by the September 2007 date.

8.3 Estimating the net impact of MA on the wider economy

8.3.1 Results by September 2001

Literature survey and simple analysis within pre-existing model of the UK economy

Advantages:
Cheap; provides basic appraisal of the possible outcomes of a more in-depth evaluation

Disadvantages:
Pre-existing models are not specifically designed for an evaluation of MA; little information available on the direct effect of MA participation on individual productivity

8.3.2 Results by January 2002

<u>Analysis within basic long-run equilibrium model of the UK economy designed for an evaluation of the wider effects of MA</u>

Advantages:
Cheap; model specifically designed for evaluating MA; provides basic appraisal of the possible outcomes of a more in-depth evaluation

Disadvantages:
Tight timetable for developing a detailed model for the evaluation; little information available on the direct effect of MA participation on individual productivity

8.3.3 Results by September 2003

<u>Analysis within fully specified long-run equilibrium model of the UK economy designed for an evaluation of the wider effects of MA</u>

Advantages:
Model specifically designed for evaluating MA; provides robust appraisal of the wider effects of MA under different scenarios for the direct effect of MA participation on individual productivity

Disadvantages:
Still relatively little information available on the direct effect of MA participation on individual productivity

8.3.4 Results by September 2005 and 2007

As for September 2003, but using increasingly reliable information on the direct effect of MA participation on individual productivity.

REFERENCES

Aghion, P. and Howitt, P. (1998), '*Endogenous Growth Theory*', MIT Press

Barrett, A., Hovels, B., den Boer, P. and Kraayvanger, G. (1998), *Exploring the returns to continuing vocational training in enterprises*. Report for CEDEFOP

Blundell, R., Dearden, L., and Meghir, C. (1996), *The Determinants and Effects of Work Related Training in Britain*, Institute for Fiscal Studies, London

Blundell, R., Dearden, L., Meghir, C. and Sianesi, B. (1999), 'Human capital investment: the returns from education and training to the individual, the firm and the economy', *Fiscal Studies*, Vol. 20(1), p. 1-23

Blundell, R. (2001), *Estimating the Returns to Education: Models, Methods and Results.* Preliminary handout prepared for the RSS meeting February 27th 2001.

Coleman, N. and Williams, J. (1998), *Evaluation of Modern Apprenticeships: 1998 Survey of Young People.* Department for Education and Employment Research Report RR93.

Cully, M., Woodland, S., O'Reilly, A. and Dix, G. (1999), *Britain at Work, As depicted by the 1998 Workplace Employee Relations Survey.* Routledge.

Dearden, L., Reed, H. and Van Reenen, J. (2000), 'Who Gains when Workers Train? Training and Corporate Productivity in a panel of British industries', IFS Working Paper No. 00/04

DfEE (1999) *Statistics of Education: Participation in Education and Training by Young People Aged 16 and 17 in each Local Area and Region, England, 1993/94 to 1997/98 (Statistical Bulletin 14/99).* London: The Stationery Office.

DfEE (2000a) *Participation in Education, Training and Employment by 16-18 Year Olds in England: 1998 and 1999.* National Statistics First Release SFR 28/2000.

DfEE (2000b) *Work-based training for young people and work-based learning for adults: Volumes and outcomes.* National Statistics First Release SFR 52/2000.

DfEE (2000c) *In their own words...20 Modern Apprentices tell their stories.*

DfEE (2001) *Modern Apprenticeships: Consultation Response,* available on-line at www.DfES.gov.uk/ma.consultation.

Dolton, P. (1993), 'The Economics of Youth Training in Britain', *Economic Journal*, Vol. 103, p.1261-1278

Dolton, P., Makepeace, G. and Treble, J. (1994), 'The wage effect of YTS: evidence from YCS, *Scottish Journal of Political Economy*, Vol. 41, p.444-453

Greenwood, J. and Yorukoglu M. (1997), '1974', *Carnegie-Rochester Conference Series on Public Policy*, Vol. 46, p. 49 – 95

Griffith, R. (1999), 'Using the ARD establishment level data to look at foreign ownership and productivity in the United Kingdom', *Economic Journal*, Vol. 109, p. F416-F442

Haskel, J. and Heden, Y. (1999), 'Computers and the Demand for Skilled Labour: Industry- and Establishment-level Panel Evidence for the UK', *Economic Journal*, Vol. 109, p. C68-C79

Haskel, J. and Martin, C. (1996), 'Skill shortages, productivity growth and wage inflation', in *Acquiring Skills: Market Failures, Their Symptoms and Policy Responses*, eds. Booth, A. and Snower, D., Cambridge University Press

Haskel, J. and Slaughter, M. (1998), 'Does the Sector Bias of Skill-Biased Technical Change Explain Changing Wage Inequality?', CEPR Discussion Paper No. 1940 and NBER Working Paper No. 6565

Heckman, J. (1979), 'Sample selection bias as a specification error'. *Econometrica,* Vol 47, p. 153-161

Heckman, J., LaLonde R and Smith, J. (1999), The economics and econometrics of active labor market programs. In *Handbook of Labor Economics* Vol. 3A (eds. Ashenfelter, O. and Card, D.), Elsevier.

Heckman, J., Lochner, L. and Taber, C. (1998), 'General-Equilibrium Treatment Effects: A Study of Tuition Policy', *American Economic Review*, Vol. 88, p. 381-386

Layard, R., Mayhew, K. and Owen, G. (1994), 'Why we need a Training Reform Act', in Britain's Training Deficit, eds. Layard, R., Mayhew, K. and Owen, G., Avebury, Aldershot

Layard, R., Nickell, S. and Jackman, R. (1991), *Unemployment: Macroeconomic Performance and the Labour Market.* Oxford University Press

La Valle, I and Shepherd, P (2000) *Feasibility Study for a Longitudinal Survey of Young People.* London: National Centre for Social Research.

Lucas, R. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, Vol. 22, p. 3-42.

Machin, S. and Van Reenen, J. (1998), 'Technology and Changes in Skill Structure: Evidence from Seven OECD Countries', *Quarterly Journal of Economics*, Vol. 113, p. 1215-1244

Manacorda, M. and Petrongolo, B. (1999), 'Skill Mismatch and Unemployment in OECD Countries', *Economica*, Vol. 66, p. 181-207

Middlemass, J. (1999) 'Modern Apprenticeships: Achievements so far'. *Labour Market Trends,* September.

Nickell, S. and Bell, B. (1995), 'The Collapse in Demand for the Unskilled and Unemployment Across the OECD', *Oxford Review of Economic Policy*, Vol. 11, p. 40-62

Nickell, S. and Bell, B. (1996), 'Changes in the Distribution of Wages and Unemployment in OECD Countries', *American Economic Review*, Vol. 86, p. 302-308

Nickell, S. and Nicolitsas, D. (2000), 'Human capital, investment and innovation: what are the connections?', in *Productivity, Innovation and Economic performance*, eds. Barrel, R., Mason, G. and O'Mahony, M., Cambridge University Press

OECD (1996), *The OECD Jobs Strategy: Enhancing the Effectiveness of Active Labour Market Policies*, OECD

Oulton, N. (1996), 'Workforce skills and export competitiveness', in *Acquiring Skills: Market Failures, Their Symptoms and Policy Responses*, eds. Booth, A. and Snower, D., Cambridge University Press

Oulton, N. (1997), 'The ABI respondents database: A new resource for industrial economics research', *Economic Trends*, No. 528, p.46-57

Payne, J. (1998) *Routes at 16: Trends and Choices in the Nineties.* Department for Education and Employment Research Report RR55.

Payne, J. (2000)*Young People not in Education, Employment or Training: Data from the England and Wales Youth Cohort Study.* Department for Education and Employment Research Report RR201.

Payne, J. (forthcoming 2001) *Work-Based Training for Young People: Data from the England and Wales Youth Cohort Study.* Department for Education and Skills Research Report.

Prais, S. (1995), 'Productivity, Education and Training: An International Perspective', NIESR Occasional Paper 48, Cambridge University Press

Riley, R. and Young, G. (1999), 'Skill Heterogeneity and Unemployment', presented at ESRC Macroeconomic seminar, Warwick University, July 1999; staff seminar, Queen Mary and Westfield college, November 1999; European Economic Association conference, Bozen - Bolzano, Italy, September 2000.

Riley, R. and Young, G. (2000), 'New Deal for Young People: Implications for Employment and the Public Finances', Employment Service Research and Development Report, ESR62, December

Skills Task Force (2000), 'Skills for all: Research Report from the National Skills Taskforce', DfEE.

Social Exclusion Unit (1999) *Bridging the Gap: New Opportunities for 16-18 Year Olds not in Education, Training or Employment.* London: The Stationery Office.

Wilkinson, D. (2001) *Individual Transitions out of Unemployment - an Evaluation of New Deal for Young People.* Interim report from the Policy Studies Institute to the Employment Service.