



BIS Evaluation Summary and Peer Review

The BIS Expert Peer Review Group for Evaluation reviews all BIS impact evaluation publications, and provides an independent assessment of the methodological quality of the evaluation.

Title: Measuring the Net Present Value of Further Education in England

Programme evaluated: Further Education

Impact Evaluation Score: 4
(see end of summary)

Monetisation Score: 4
(see end of summary)

Time period covered by policy: On-going	Time period covered by evaluation: Data covering learning in FE between 2002/3 – 2010/11 academic years
Contractor undertaking evaluation: BIS analysis	Peer reviewers: Dr Edward Anderson Dr Steven McIntosh Professor Anna Vignoles
Type of evaluation: Outcome evaluation (score 4) – the report uses estimates from a study (Urwin 2014) which estimates the wage and earnings premia for different types of Vocational Education provision and incorporates these into a model (Cambridge Econometrics, 2011) which converts these premia and other data and assumptions into Net Present Values.	

Description of policy/programme and rationale for intervention:

Adult skills funding seeks to address market failures which result in individuals and employers failing to invest in the skills required to support the economy. This can be used to help individuals continue their education, where they may have failed at school; to support unemployed individuals who need retraining to help them go back into the labour market and to support employers in increasing skill levels of their current and future workforce and hence their productivity

Summary of key evaluation findings:

1. BIS have updated their estimates of the NPV of VE qualifications.
2. Based on these updated estimates, publically-funded qualifications in the Further Education sector in England continue to show strong economic returns (£20 per £ of funding, gross of deadweight and £14 per £ of funding net of deadweight)

Summary of cost-benefit/cost-effectiveness analysis (if applicable):

As above.

Policy response to the evaluation:

Policy colleagues have noted the reduction in NPV per £ since the previous estimates and welcome the increased robustness of the employment premia, the principle cause of the reduction.

Evaluation methodology

Description of methodology:

The NPV estimates use a model of the impact of FE (BIS Research Paper 38¹), and estimates from a study (Urwin 2014²) using matched administrative data on wage, employment and benefit claims, before and after learning, to compare labour market outcomes for achievers of a particular highest learning aim against those who have the same highest learning aim, but do not achieve it.

Does the evaluation review the published policy objectives?

N/A

At what level are the main intended outputs and/or outcomes expected to occur? (What is the unit of analysis? For example: universities, businesses, individuals or nationally)

Individuals

Has sufficient time lapsed for the initial/full benefits to be estimated?

Yes – labour market outcomes are observed from 1-7 years after completion of learning.

Peer review

Comments on the appropriateness of data and outcomes:

Professor Anna Vignoles:

The underlying estimates of the impact of FE qualifications on wages are based on high quality administrative data. The earnings data in particular is likely to be higher quality than in many other studies given that they come from tax records. However, other aspects of the model are based on data that may be more problematic, such as the data on guided hours. The extent to which guided hours are correlated with actual hours of study is unknown and hence estimates of the costs of study may

¹ **Cambridge Econometrics (2011)** Measuring the Economic Impact of Further Education, March 2011 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32329/11-816-measuring-economic-impact-further-education.pdf

² **Urwin (2014)** Further education: comparing labour market economic benefits from qualifications gained, December 2014. <https://www.gov.uk/government/publications/further-education-comparing-labour-market-economic-benefits-from-qualifications-gained>

have error though the authors note that the benefits are sufficient that they are the main driver of the NPV.

Dr Edward Anderson:

Overall the data and outcomes referred to in the report are appropriate and follow standard practice in the literature. The report aims to measure the main economic benefits of further education (FE), but not to attempt to measure the wider social benefits such as improved health, reduced crime, or higher civic engagement. The economic benefits of FE are made up of higher productivity and/or employability of those graduating from further education, plus spillover benefits of FE to other workers and/or firms in the economy. The costs of FE are made up of the costs of provision (partly paid for by the government, and partly paid for by individuals or firms), plus the opportunity cost of the time taken by learners to undertake training. This matches up with standard practice in the literature.

Dr Steven McIntosh:

The analysis has taken an initial model developed by Cambridge Econometrics and updated it with new estimates of wage returns to various vocational qualifications, employment probabilities and deadweight. These are the most recent estimates available in the literature, and so are the most appropriate to use. It is important to make clear, as the report does, that these estimates are produced using a very different methodology to those used previously. The new estimates (from Urwin, 2014) are based on analysis of administrative data for participants observed in Further Education, and then matched to tax and benefit data to obtain wage information. Previous wage returns analysis has used survey data, comparing individuals whose highest qualification is at one level, to those at the next lowest level. The new administrative data, containing as it does only learners, allows the comparison of those individuals who complete a learning aim with those who started, but failed to complete, the same learning aim.

There are various advantages to using the estimated wage returns and employment probabilities from Urwin (2014) based on the administrative data. The analysis is based on extremely large numbers of observations (basically, the population of learners), with very detailed information on learning aims and achievement/non-achievement, and tracking of the learners into the labour market through the matched data. Thus appropriate data for the estimated returns to feed into the model have been used.

The other main source of new information that has been fed into the original model concerns estimates of deadweight, that were not included at all when the model was first used due to a lack of suitable information. Since the 2011 Cambridge Econometrics report, estimates of deadweight for apprenticeships have been produced by London Economics (2012). These estimates have been applied to all vocational qualifications in this current updating of the model. While it is possible to think of reasons why the extent of deadweight could differ across qualifications (different levels of funding, different employer attitudes to providing training for different qualifications in the absence of funding), it seems appropriate to me to use what information on deadweight is available, rather than ignore the issue altogether. The report also helpfully provides a sensitivity analysis of the deadweight

assumption. Obtaining estimates of deadweight for other qualifications would seem to be a fruitful area for future research, though, the difficulty of doing so notwithstanding.

Comments on internal validity:

Professor Anna Vignoles:

One threat to internal validity arises in the calculation of the underlying estimates of the impact of FE qualifications on earnings and employment. The impact estimates are based on a comparison of the earnings of those who complete a particular FE qualification and those who start but do not complete the qualification. This comparison may not be credible if qualifications act as a filter of ability so that non achievers are less able than achievers. One might also argue that non achievers may have weaker motivation, again leading to bias. An alternative explanation is that non achievers may have simply got a job and that is why they did not complete their qualification. In this latter case, the bias may go the other way. This only appears to be a potential issue for the calculations for a full Level 2 qualification however, since for other qualifications previous estimates of the wage return based on different approaches yield similar estimates using this new approach. For a full level 2 previous estimates suggested a return of just 2% whilst this new method yields returns of 11%.

The estimates focus on highest qualification achieved and hence do not estimate the value of lower level qualifications that enable progression to higher level qualifications. The value of lower level qualifications may be understated.

Dr Steven McIntosh:

The internal validity of the estimated NPV figures presented here depends on the internal validity of the labour market returns data that are fed into the model. As mentioned in the previous section, these returns are from the Urwin (2014) report, which compares achievers to non-achievers with the same learning aim. Clearly this is a non-experimental setting, and the possibility exists for characteristics to differ between the treatment and control groups, which then have a confounding effect on the wage and employment differences between them. Note that there is no way to determine whether the effect of such confounding factors is greater than or less than the effect of similar confounding factors when learners are compared to non-learners at a particular level, as typically used in such research and as used in the original version of the model presented here.

The report helpfully provides estimates of wage returns and employment probabilities estimated via the 'traditional' method as used in the original version of this model, and those of Urwin (2014) estimated with the new methodology based on administrative data. It is pleasing that for most qualifications, the differences in estimates are not very large, with in most cases the new methodology with the non-achieving comparison group providing slightly lower estimates. The key exception, in terms of both the difference in estimates being marked and the new methodology providing the larger estimate, is for full level 2 qualifications. I tend to agree with the report here, that the comparison of level 2 learners to non-learners (i.e. those with low or no qualifications) can involve comparing very different types of individuals,

given that the latter are a very heterogeneous group who have no or only low qualifications for a wide variety of reasons. The non-achieving level 2 learners are therefore likely to form a better comparison group, formed as they are of 'the type of individuals' who enrol for a level 2 vocational qualification.

Nevertheless, as stated above, given that the treatment group have achieved their learning aim whilst the comparison group have failed for some reason, differences in the individual characteristics between the two groups at least potentially still exist. It is therefore important to mitigate the influence of any such confounding factors as far as possible, which the source of the estimated returns (Urwin, 2014) does as described in the causation section below.

Comments on external validity:

Professor Anna Vignoles:

Labour market outcomes are observed from between one and seven years after the episode of learning. This is a relatively short time period. The model assumes that beyond the period of the data those who achieve the qualifications have parallel earnings trajectories to those who do not but this may not be the case. Hence the calculations do not necessarily estimate the lifetime benefits of the FE qualifications.

Dr Steven McIntosh:

The NPV results derived in the current report use the estimated labour market returns obtained through an analysis of the population of learners, as described above. As such the obtained NPV results are representative for this group of learners. Readers should therefore be aware that the results cannot be generalised to the whole population – such NPV figures would not be achieved by any individual undertaking vocational education, and for many, vocational education will add little in value to their existing education (with the added necessary consideration of general equilibrium factors if vocational education was to be significantly expanded). It seems to me that the results in the current report are exactly the ones we want though – we want to know the value of obtaining these qualifications to the individuals who would want to obtain them, rather than averaged across the full population, many of whom have not and will never obtain such qualifications.

Another issue to consider here is that the estimated returns used in the update to the model only observe individuals for up to 6 years after the completion of their learning, rather than being observed across all ages. This is a necessity due to nature of the available data. As time progresses and the matched administrative data set expands, it will clearly be of interest to obtain estimated returns over a longer time period after learning. Until such time, though, the approach taken in the current report, to assume that the estimated returns are persistent over the working lives of the individuals, seems reasonable, given the evidence presented in Figures 1-4 in Annex 1 of the report, which show returns maintaining a flat profile over the time period for which the learners are observed in the matched data set.

Comments on the quality of inferences and establishing causation:

Professor Anna Vignoles:

Overall the analysis rests on estimates of wage and employment effects that have some strengths (good data) and some weaknesses (a non-experimental control group). The caveats and assumptions behind the model are carefully explained and some sensitivity analysis is included.

Dr Edward Anderson:

The most important causal links in calculating the NPV of further education are the effects of further education on wage and employment levels. Estimates of these effects – referred to as ‘wage and employment premia’ – are taken from a separate study (Urwin 2014). Although I have not examined this other study, the overall impression I get is that the estimates are based on a sound methodology and plausible assumptions, e.g. regarding the persistence of wage premia over time.

It is good that the report takes into account the potential spillover benefits of further education, even though these are difficult to estimate.

Dr Steven McIntosh:

As commented in the internal validity section, the extent to which a causal interpretation can be put on the results, that the observed NPVs are due to the vocational qualifications, and not to any other confounding factors, depends upon the extent to which such confounding factors were controlled for in the original derivation of the labour market returns that are fed into the model here. The source of those returns (Urwin, 2014) went to great lengths to control as far as possible for any potential differences between achievers and non-achievers at the same level. The first point to note, and the key benefit of using the matched administrative data set, is that both groups comprise individuals who chose to enrol in and study for vocational qualifications at the same level. The analysis can therefore hold constant the motivation to study between the two groups, something that ‘traditional’ analyses have found difficult. A long list of control variables are also included in the analysis, to control for other differences between the two groups. The full list is provided in footnote 1 in the report. Key amongst this list are variables measuring the number of previous learning spells, the number of days on benefits prior to learning and the number of days in employment prior to learning, allowing the authors to control for the previous history of the two groups prior to the learning spell in a way not usually possible to researchers using survey data, and so ensure that the achievers and non-achievers to be compared were as similar as possible going into their learning spell.

Any other comments:

Dr Steven McIntosh:

I would just add that the analysis in the current report shows the benefits of setting up the original model (Cambridge Econometrics, 2011) in such a way that new data on returns etc. can be slotted easily into the model to produce updated estimates of the NPV of qualifications.

Cost-effectiveness and cost-benefit summary

Justification for monetisation score:

The NPV estimates use monetised benefits and costs of Further Education qualifications, taking deadweight into consideration.

Sensitivity analysis/key assumptions:

The model of the impact of FE (BIS Research Paper 38¹) was published before the introduction of the BIS peer review framework. The new estimates covered in the paper use the original model framework (updated with better and more up-to-date evidence). The peer review covers the new NPV estimates and the elements of the 'Impact of FE' model that were used to derive the new estimates.

Aspects of the original model not included in the peer review are:

- 1) The original wage and employment premia because these have been updated with better more up-to-date evidence
- 2) Costs and benefits to the exchequer because the exchequer side of the model has never been used BIS so is excluded from the peer review. If BIS find occasion to use that side of the model it would be peer reviewed at that stage.

The key assumptions for the new NPV estimates are:

1. **3-5 year (average) wage returns**
2. **3-5 year (average) employment returns**
3. **Persistence of the 3-5 year average wage and employment premia over the lifecycle**
4. **Parallel age-earning base-profile for achievers and non-achievers (may under-state the benefits)**
5. **Application of 30% deadweight (calculated for apprenticeships) to other provision types (in some cases this will under-state the benefits and in others it may over-state the benefit)**

Sensitivity analysis and a full list of key assumptions for the original 'Impact of FE' model (BIS Research Paper 38¹) are presented in Annex 2 of the publication. The key assumptions are reported here

6. **100% spill-over benefit**
7. **Steady state of the labour market for intermediate level skills in the future**

Description and size of key monetised costs:

- Forgone output during training is calculated by multiplying the guided learning hours associated with each qualification, by the national average wage at their previous highest qualification level.
- Participation funding is taken from the SFA ILR data used for the annual provider funding allocations.
- Fees paid by companies or individuals

Other key non-monetised costs: n/a

Description and size of key monetised benefits:

Benefits are based on the impact of qualifications on Gross Value Added:

- Wage effects are calculated by applying wage premia to the national average wage at their previous highest qualification level.
- Employment premia account for the wages of those who would otherwise have been unemployed or more prone to spells of unemployment.
- Spillovers to others including increases in other worker's wages and productivity, e.g. from knowledge sharing; increases in firm productivity, e.g. better use of current capital and increases in other firms productivity, e.g. competing through reciprocal human capital investment.

Other key non-monetised benefits:

- 'Social' benefits, such as reduced crime or improved health

Robustness of monetised costs and benefits:

Overall we judge the monetization to be robust because it is based on peer reviewed wage and employment returns. We comment on other key assumption in detail within the paper. The original model itself relies on assumptions that we judge to be robust after conducting sensitivity analysis.

Peer Review

Evaluation peer review comments on comprehensiveness, clarity, robustness and best practice of cost benefit/cost effectiveness analysis:

Professor Anna Vignoles:

The model to value costs and benefits is complex and provides quite disaggregated results which are based on a lot of different assumptions. Whilst some sensitivity analysis is conducted, not every assumption is tested. The model does include key costs and provides clear reasoning behind the assumptions.

The model does not include a range of potentially economically valuable outcomes which will affect the CBA. For example it does not include private benefits such as health or wellbeing gains. For example it does not include social benefits such as reduced crime.

The model does not include a range of costs, including some social costs of FE such as leisure time foregone to undertake study.

The calculation of the NPV of the full Level 3 qualification depends on assumptions about the RAB charge which may be lower or higher than assumed thereby altering this cost. The assumptions in the report are entirely consistent with government calculations of the RAB charge however.

Dr Edward Anderson:

Overall the cost benefit analysis is carried out well and follows standard practice in

the literature. The revised report has been strengthened by including estimates of the IRR as well as the NPV of further education in the UK, and by carrying out a more extended sensitivity analysis.

Dr Steven McIntosh:

Having commented on the appropriateness and validity of the source information on labour market returns in the earlier peer review section, this section discusses their use in the estimation of the NPV.

Overall, the NPV is comprehensive in terms of the costs and benefits included, certainly comparable with other NPV estimates in the literature, and follows best practice. The key benefits are the economic gains to learners in terms of higher wages and employment probabilities, which are monetised by applying the estimated labour market returns to the actual observed wages and employment rates of those individuals at each age with the immediately lower qualification level. Consistent wage and employment premia are employed at every age, given the absence of available information on how the premia vary with age. This seems reasonable, particularly given the evidence presented in Figures 1-4 in the report in Annex 1 which show constant estimated premia as age increases, over the time period for which learners are observed in the matched data set.

The other key economic benefit included is spillover effects, whereby the acquisition of the qualification by an individual can also improve the productivity of co-workers. The report for the original construction of the model (Cambridge Econometrics, 2011) makes clear that such spillover effects are assumed to be equal in size to the improved productivity of the learner, so that the latter is essentially doubled when assessing the total benefits. This assumption is based upon the earlier analysis of Dearden et al (2005) whose analysis at the industry level showed that the productivity effects of training are double the wage effects, with the difference assumed to be due to spillovers. The use of such an assumption in the current report therefore seems reasonable in the absence of additional information on spillovers. Obtaining such new information would be a useful research avenue to pursue. For now, the current report includes a helpful sensitivity analysis of the spillover assumption.

The principal omissions from the list of costs and benefits, in common with other estimates of economic NPV available in the literature, were the 'wider' benefits, such as improved health, reduced crime, increased societal involvement etc., or the 'wider' costs of reduced leisure and caring time. Given the difficulty of placing a monetary value on such costs and benefits, this seems to be appropriate.

The full NPV model is based on a number of key assumptions, that always have to be made in such exercises. The assumptions are made primarily to keep the model tractable (for example, assuming no general equilibrium effects such as the effects in other areas of raising taxes to provide the funding, or the effects on the employment of other individuals). The assumptions made all seem reasonable, and on the margins should not influence greatly the estimated NPV of additional learners.

Note on Impact Evaluation and Monetisation Scores

Impact Evaluation Score

Impact scale follows the guidance on 'Quality on Impact Evaluation'³, published as supplementary guidance to the Magenta Book. The scale is based largely on the Maryland Scientific Method Scale used by academics and researchers to assess the strength of an evaluation approach. The higher the score potentially the more capable the evaluations are to demonstrate that the outcome observed is due to or caused by the intervention.

- Score 5: Random allocation of treatment and control group, or a robust counterfactual using a quasi-experimental approach. There is a treatment and a comparison group and actual before and after data in both groups. For example: a strong difference-in-difference design, regression discontinuity design or matched treatment and control group.
- Score 4: Quasi-experimental approach where the counterfactual has some weaknesses, but it is as good as can be, given the policy design or data availability issues. There is a treatment and a comparison group, and actual before and after data in both groups. For example: a difference-in-difference design, regression discontinuity design or matched treatment and control group.
- Score 3: Predicted (modelled) versus actual outcomes for the treatment group only are compared, predictions are based on actual baseline data.
- Score 2: Actual (i.e. not self-assessed or self-reported) before and after data for the treatment group only are compared. (Higher levels on this scale also require actual data not based on self-reporting.)
- Score 1: No baseline data (or only self-assessed/self-reported data).

Monetisation Score

The higher the score the more information the evaluation contains in terms of analysing the cost of the intervention and the additional benefits to the economy.

- Score 5: Input, output, outcome data additional Benefit Cost Ratio (BCR), NPV set aside some other not monetised impact measures, fuller cost benefit analysis or cost effectiveness analysis that compares the costs of alternative

³ Quality in policy impact evaluation, HMT, Dec 2012

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation_QPIE_.pdf

ways of producing the same or similar outputs

- Score 4: Input, output, outcome data, calculation of additional Benefit Cost Ratio, Net Present Value
- Score 3: Input, output, outcome data calculation of Gross BCR not additional or not clear if additional
- Score 2: Gross BCR not available, as either input or output data are not available
- Score 1: No monetisation at all

BIS/15/326