Contents

Preface 3
Executive Summary 4
Introduction 9
Impact evaluation 13
Methodology 16
Definition 19
Findings 23
Summary of findings 33
References 36
Appendix A: Findings by outcome 38
Appendix B: Evidence Reviewed 39
Preface

This report presents findings from a systematic review of evaluations of the impact of apprenticeships on firms and workers. It is the eighth review produced by the What Works Centre for Local Economic Growth. The What Works Centre is a collaboration between the London School of Economics and Political Science, Centre for Cities and Arup and is funded by the Economic & Social Research Council, The Department for Communities and Local Government and The Department for Business Innovation & Skills.

These reviews consider a specific type of evidence – impact evaluation – that seeks to understand the causal effect of policy interventions and to establish their cost-effectiveness. To put it another way they ask ‘did the policy work’ and ‘did it represent good value for money’? By looking at the details of the policies evaluated we can also assess what the evidence tells us about delivery issues – for example, is there any evidence that schemes with a particular sectoral focus do better than other schemes?.

Process evaluation – looking in detail at how programmes operate – provides a valuable complement to impact evaluation, but we do not focus on this. We recognise that may sometimes cause frustration for practitioners who are responsible for delivery.

We see these impact-focused reviews as an essential part of more effective policy making. We often simply do not know the answers to many of the questions that might reasonably be asked when implementing a new policy – not least, does it work? Figuring out what we do know allows us to make better decisions and to start filling the gaps in our knowledge. This also helps us to have more informed discussions and to improve policy making.

These reviews therefore represent a first step in improving our understanding of what works for local economic growth. In the months ahead, we will be working with local decision makers and practitioners, using these findings to help them generate better policy.

Henry Overman;
Director, What Works Centre for Local Economic Growth
Executive Summary

This report presents findings from a systematic review of evaluations of the economic impact of apprenticeships, focusing in particular on the impact on workers and firms. It is the eighth review produced by the What Works Centre for Local Economic Growth.

The review considered more than 1,250 policy evaluations and evidence reviews from the UK and other OECD countries. It found 27 impact evaluations that met the Centre’s minimum standards.

Approach

The Centre seeks to establish causal impact – an estimate of the difference that can be expected between the outcome for workers undertaking, or firms offering, apprenticeships and the average outcome they would have experienced without the apprenticeship (see Figure 1). Our methodology for producing our reviews is outlined in Figure 2.

Figure 1: Evaluating impact

Change in outcome for those on the programme

VS

Change in outcome for those not on the programme
Findings

This review considers the impacts of apprenticeships on workers and firms. This section summarises the detailed findings. We emphasise that many of these findings depend on a small number of studies. They are, however, consistent with other research on apprenticeships and on employment training more broadly (see our first evidence review).

What the evidence shows

- There is some evidence that apprenticeships improve skill levels, and stimulate further training or study.
- Apprenticeships can increase wages, although in two evaluations effects are negative. Impacts also vary by type of participant.
- Apprenticeships tend to have a positive effect on participants’ subsequent employment (and also reduce subsequent unemployment).
- Level 3 or higher apprenticeships deliver substantially higher lifetime wage gains relative to lower level apprenticeships (based on the limited UK evidence available).
- There is some evidence that apprenticeships are more likely to increase employment than other forms of employment training (unless that training also involves an in-firm element). The evidence of impact on wages is more mixed and appears to vary by gender.
- There is some evidence that identifies mechanisms that may increase entry into apprenticeships and attendance during the programme (e.g. pre-qualifications, higher wages and subsidies to individuals). However, we have less evidence on what works to ensure people complete apprenticeships.
Where the evidence is unclear

- It is unclear whether the duration of the apprenticeship matters for effects on wages or employment (although longer apprenticeships that deliver higher qualifications may have more positive effects)

Where there is a lack of evidence

- There is some limited evidence that firms participating in apprenticeships experience economic gains, such as higher productivity or profits. This fits with survey evidence, but more impact evaluations are needed.
- There is too little evaluation evidence to draw clear conclusions on whether apprenticeships work better in some sectors than others.
- There is some evidence that post-apprenticeship moves can increase wages although effects depend on circumstances.
- There is no impact evaluation evidence looking at the effect of apprenticeships on a given local area (rather than individual participants or firms).
- There is no impact evaluation evidence comparing the effects of nationally run programmes versus locally run programmes.
- Existing ex-ante modelling suggests that the economic benefits of apprenticeships comfortably outweigh their costs. However, only one of the impact evaluations provides cost data in a form which allows us to calculate ex-post benefit-cost ratios for that programme.
- None of the shortlisted studies look at the effects of substantially scaling up apprenticeship provision, as is currently happening in the UK. We need more evidence on whether identified benefits also hold in a larger programmes. Given the other substantial changes to the UK apprenticeship system in the past decade and a half, more up to date UK impact evaluation evidence is also needed.

How to use these reviews

Apprenticeships are currently very high on the policy agenda, and the evidence review highlights a number of factors for policy makers to be aware of when considering apprenticeships:

- While the evidence suggests that higher level apprenticeships (specifically, Level 3 and above) may offer better outcomes, it does not currently tell us whether this is because stronger candidates gravitate towards more demanding programmes. If this is the case, policymakers need to consider how to address the needs of those ‘left behind’ by this type of apprenticeship offering.
- Any policy should carefully consider how to recruit firms to provide apprenticeships, and trainees to fill them. A better understanding of the costs and benefits to firms will help in this (see below), as will a better understanding of which policy design aspects increase take-up and reduce drop-out.

To determine policy priorities

The Centre’s reviews consider a specific type of evidence – impact evaluation – that seeks to understand the causal effect of policy interventions and to establish their cost-effectiveness. In the longer term, the Centre will produce a range of evidence reviews that will help local decision makers
decide the broad policy areas on which to spend limited resources. Figure 3 illustrates how the reviews relate to the other work streams of the Centre.

Figure 3: What Works Centre work programme

Helping to fill the evidence gaps

As should be clear from this review, there are many things that we do not know about the local economic impact of apprenticeships.

If achieving local economic impact is an important part of the case for apprenticeship provision, then there need to be more evaluations that explicitly explore these impacts and how to maximize them. Central and local policymakers – and private sector partners – should:

- Look to undertake systematic comparisons that cover different kinds of apprenticeship model - for example, the German system versus a more decentralised system.
- Conduct further research looking at outcomes for firms. Surveys of firms who offer apprenticeships suggest those firms see clear benefits, but they may not be representative of all employers.
- Set up evaluations of scheme design and its effect on take-up, completion and outcomes. This is particularly important given devolution of skills budgets to cities such as London and Manchester. Central and local policymakers should work together to design robust evaluation that increases our understanding of how to improve the design of apprenticeships.
- Make scheme cost data available to researchers so that robust benefit-cost ratios can be calculated.

The Centre’s longer term objectives are to ensure that robust evidence is embedded in the development of policy, that these polices are effectively evaluated and that feedback is used to improve them. To achieve these objectives we want to:

- Work with local decision makers to improve evaluation standards so that we can learn more about what policies work, where.
- Set up a series of ‘demonstration projects’ to show how effective evaluation can work in practice.

Interested policymakers please get in touch.
Introduction

This review looks at the impacts of apprenticeship programmes on workers and firms. Apprenticeships involve paid employment in a firm, alongside training provided by that employer (and sometimes others), typically leading to a formal qualification or title:

*Apprenticeships involve a program of courses, work-based learning, and productive employment in which workers achieve occupational mastery and industry-recognized credentials.* Unlike school-based vocational education, apprenticeships involve extensive work-based learning and practice; real jobs involving production, pay, and the discipline of work; and close mentoring by professionals. Unlike on-the-job training contracts, apprenticeships include related courses and the development of occupational mastery, not simply the ability to do a particular job. [1] [our emphasis]

As the 2012 Richard Review points out, 'occupational mastery' is crucial: apprenticeships involve participants learning a job role that is new to them and requires substantial study and practical commitment. [2]

The use of apprenticeships by employers dates back hundreds of years. More recently, the perception that apprenticeships are also effective in improving labour market outcomes have made apprenticeships increasingly popular with policymakers and employers. In the UK, for example, government has run apprenticeship programmes since 1994. [3] The UK has historically followed a market-based approach, in which firms take the lead; apprenticeships are a form of temporary employment contract, lasting between one and four years, and there is no formal link to the education system.

Following the 2012 Richard Review, the previous system, which involved over 250 different apprenticeship frameworks, is now being simplified and formalized, with some graduate-level apprenticeships introduced that are designed to offer a formal alternative to a full-time degree. [4] The number of apprenticeships is also increasing. For example, since 2010, the number in England has doubled, with accompanying increases in funding; however, there are still fewer apprentices as a

1 Lerman (2014).
2 This is the definition adopted by the 2012 Richard Review.
3 CEBR (2013). English, Welsh and Scottish governments run their own programmes.
share of the workforce in England (11 per 1,000 employees) than some other countries (39/1,000 in Australia, 40/1,000 in Germany and 43/1,000 in Switzerland).\(^5\)

The Government is committed to further increasing the number of apprenticeships. The 2015 Summer Budget announced the creation of 3 million new apprenticeships by 2020, funded by a levy on large employers. The exact details of this levy, including rates and implementation, have yet to be established but firms will have access to the funds raised to cover the costs of post-16 Apprenticeships in England. Money to pay for the training will continue to go straight to providers rather than employers, with employers accessing the funding via an apprenticeship voucher system. The overall aim is to increase employer investment in training.

In addition to these national level changes, devolution will also have implications for policy, with Greater Manchester now having control over the employer apprenticeship grants budget, and other cities potentially following suit in the future.\(^6\)

This review focuses on the economic impacts of apprenticeship programmes on participating individuals and firms (and by implication, the local areas in which they are based).

Large claims are sometimes made for these economic impacts.\(^7\) For example, a recent BIS report suggests that the net present value of Level 2 and Level 3 apprenticeships begun in 2013/14 are £12bn and £10bn respectively.\(^8\) These are substantial numbers. In addition to helping with programme design, the impact evaluation evidence reviewed in this report is one element in determining whether or not such predictions are plausible.

What can we expect apprenticeship programmes to achieve? There is a large existing literature on the economic returns to training and education, including some studies that look directly at apprenticeships. These evaluations typically find large, significant positive wage and employment gains to individuals who have participated in apprenticeships, versus individuals who did not.\(^9\) Not all of these studies meet our quality thresholds, but we include any suitably robust evaluations in our shortlist.

What might drive these effects? Apprenticeships aim to improve individuals’ employment and wage outcomes by raising their human capital. Apprentices should be more employable after the programme, and have mastered specific skilled roles. This should raise lifetime wages, reduce the risk of future unemployment and improve prospects of career progression.

Apprenticeships may also help employers, by providing them with more skilled (and thus more productive) employees. Apprentices may also be more loyal and contribute more effort than regular entry-level workers. Of course, after completing the programme apprentices may be poached by other employers, who then reap the benefits. For this reason firms will tend to underprovide apprenticeships, which is why governments provide support to firms to cover some of the cost of the programme (through grants, wage subsidies or some combination of these). These policy features have the additional benefit (to employers) of making apprentices cheaper than other, equivalently qualified, members of their workforce.\(^10\)

As explained further below, figuring out the ‘causal’ impact of apprenticeships on firms and workers is not straightforward. Issues include agencies or firms ‘cherry picking’ the most able apprentices;

---

5 Todd (2014).
6 HM Treasury and GMCA (2014).
8 BIS (2015).
9 See McIntosh and Garrett (2009) for a recent review.
10 Picchio and Staffolani (2013).
the ‘best’ or ‘worst’ firms choosing to participate in a programme; identifying the reasons for post-apprenticeship moves in the labour market, since the most able apprentices may be the most likely to move; and tracking participants (and control group members) over time, particularly for studies that seek to pick out the long term effects of apprenticeship programmes.

Figuring out the impacts of apprenticeships on a local economy is even harder. Even if we know the individual benefits, worker (and firm) location decisions affect the geographical distribution of those benefits. For example, to the extent that apprentices in a given area stay in that area after completing the programme, workforce quality in the area rises; however, if participants move to another area that human capital is lost to the new location. Of course, human capital across the UK rises either way.

These complexities also make it more difficult to estimate the net benefits of apprenticeship schemes, even if cost data is available (which is only the case in one of our shortlisted studies).

Apprenticeship systems across countries

A final complication arises from the fact that not all apprenticeship systems are the same. A lot of discussion in the UK focuses on the German system, in which apprenticeships are an integral part of the national education system, with centralized provision and chambers of commerce closely involved in regulating content and quality. Over 40% of general secondary school leavers in Germany complete an apprenticeship, which can last up to three years.11 Other European countries such as Switzerland, Austria, Norway and Denmark also follow this approach. Countries such as the US take a hybrid approach, with some centralized programmes, but these are very small and focused on sectors such as construction or manufacturing.12 Other programmes are devolved and designed by individual states, or through public-private partnerships.

As a result of this complexity, it is unsurprising that the evaluations that we end up considering cover countries that approach apprenticeships in a slightly different way. As noted, this makes them both harder to define and more difficult to compare. That said, the programmes we review can be grouped into three overarching categories: the dual educational system, employment contracts and a combination approach.

Dual Educational System

This approach is particularly strong in Germany, and has been followed by some of its European neighbours, including Switzerland, Austria, Norway and Denmark.13 The key feature of this model is centrally co-ordinated, government-led provision, with apprenticeships formalised as part of the national education system (Cooke, 2003 – Paper 847). Provision is typically facilitated through a ‘dual educational system’ which offers different schooling routes for young people: the academic route, typically leading to university and higher education (Matura in Austria; Gymnasium in Germany) and the vocational track, which comprises compulsory vocational training and leads to apprenticeship (Hauptschule in Germany and Austria), although this format varies from country to country (Cooke, 2003 - Papers 847; Fersterer et al., 2004 - Paper 878). Apprentices divide their time between on-the-job training (65-70%) and educational training (1 to 2 days a week), provided typically by the state (Parey, 2009 - Paper 986). In this model, the government is normally responsible for covering the cost of all apprentices’ classroom training. A key element of this apprenticeship model is the formalised,

---

11 REF - TBC
12 REF - TBC
13 Lerman (2014).
institutionalised nature of apprenticeships, and the regard with which apprenticeships are held nationally; as a mark of the popularity of the apprenticeship route in these countries, over 40% of general secondary school leavers in Austria complete an apprenticeship (Fersterer et al., 2004 - Paper 878).

**Employment Contract**

The approach in other European countries, such as the UK, is very different from that outlined above. Rather than a government-led system, the onus is on the private sector to provide apprenticeships.\(^\text{14}\) In these systems, the apprenticeship is seen more as a form of temporary employment contract, which allows the firms to pay less social contributions, and in return asks that they provide apprentices with certified on-the-job and classroom training (Capellari et al., 2012 – Paper 897). This is traditionally not linked to the education system; a student may choose to leave school and commence an apprenticeship, but this is not a formal track set by government (ibid). There is no centrally controlled apprenticeship programme and as such there time split between work and training varies from apprenticeship to apprenticeship. Training is often provided in firm, but external courses can also be used – this could link to courses provided by government education bodies (ibid). Funding for training is typically split between government and businesses in this model; in the UK, Government only contributes 50% of the cost of training for apprentices aged 19 – 24 (Skills Funding Agency, 2015)\(^\text{15}\).

**Combination approach**

Countries such as the US – which has historically had a very small apprenticeship system with little coverage outside construction and manufacturing – operate a hybrid approach.\(^\text{16}\) The US has a growing apprenticeship system, with President Obama announcing a new $100 million grant program to support the development of innovative apprenticeship programs across the country in 2015 (White House Office of the Press Secretary, 2015)\(^\text{17}\). Centralised apprenticeship programmes exist, which are government-led, such as the Registered Apprenticeship which is administered by the Office of Apprenticeship and State Apprenticeship Agencies at a national level. In addition, the US have implemented School to Work reforms which focus on high school students, with some elements of the academic / vocational tracking seen in Germany (Neumark and Rothstein, 2005 – Paper 964). In addition, however, there is also an emphasis on facilitating more private sector provision of apprenticeships. The US department of labour runs an ‘American Apprenticeship Grant’ which invites public-private partnerships to implement apprenticeship systems (US Department of Labour, 2015)\(^\text{18}\). Programmes for apprentices are therefore organised on a state by state basis by private, public and third sector bodies, as well as at a national government level. (Veum, 1995 – Paper 976).

**Cross-country comparisons of the evaluation evidence**

Clearly, this complexity creates challenges when comparing the results of evaluations of schemes in different countries. As with our other evidence reviews, we still think that it is possible to further our understanding of apprenticeships by looking at, and comparing, the available impact evaluation evidence from across the OECD. But this discussion highlights some of the caveats that must apply to the findings emerging from our review. Where necessary, we highlight specific issues further, below.

\(^\text{14}\) Lerman (2014).
\(^\text{16}\) Lerman (2014).
Impact evaluation

Governments around the world increasingly have strong systems to monitor policy inputs (such as spending on public subsidies for apprenticeships) and outputs (such as the total number of apprentices trained). However, they are less good at identifying policy outcomes (such as the impacts of apprenticeships on participating individuals or firms). In particular, many government-sponsored evaluations that look at outcomes do not use credible strategies to assess the causal impact of apprenticeship policies (henceforth, we refer to these as ‘programmes’).

By causal impact, the evaluation literature means an estimate of the difference that can be expected between the outcome for individuals, firms or areas implementing a programme (in this case, taking part in an apprenticeship scheme) and the average outcome they would have experienced without the project. Pinning down causality is a crucially important part of impact evaluation. Estimates of the benefits of a project are of limited use to policy makers unless those benefits can be attributed, with a reasonable degree of certainty, to that project.

The credibility with which evaluations establish causality is the criterion on which this review assesses the literature.

Using Counterfactuals

Establishing causality requires the construction of a valid counterfactual – i.e. what would have happened to an area (or part of an area) if the project hadn’t happened. That outcome is fundamentally unobservable, so researchers spend a great deal of time trying to rebuild it. The way in which this counterfactual is (re)constructed is the key element of impact evaluation design.

A standard approach is to create a counterfactual group of similar places not undertaking the kind of project being evaluated. Changes in outcomes can then be compared between the ‘treatment group’ (people or firms participating in apprenticeships) and the ‘control group’ (people or firms not participating).

A key issue in creating the counterfactual group is dealing with the ‘selection into treatment’ problem. Selection into treatment occurs when people or firms participating in apprenticeships differ from those who do not do so.
An example of this issue for apprenticeships would be when the agency implementing an apprenticeship programme selects the most capable employers and/or apprentices for the programme (‘cream-skimming’). If this happens, estimates of policy impact may be biased upwards because we incorrectly attribute better economic outcomes to the programme, rather than to the fact that the participants are already likely to perform well.

Selection problems may also lead to downward bias. For example, if a local authority programme explicitly targets the ‘hardest to help’ for a programme – such as the very long term unemployed, or those with chaotic lifestyles, then we may mistakenly attribute poor economic outcomes to the programme rather than the underlying challenges facing those taking part.

These factors are often unobservable to researchers. So the challenge for good programme evaluation is to deal with these issues, and to demonstrate that the control group is plausible. If the construction of plausible counterfactuals is central to good policy evaluation, then the crucial question becomes: how do we design counterfactuals? Box 1 provides some examples.

---

**Box 1: Impact evaluation techniques**

One way to identify causal impacts of a project is to randomly assign participants to treatment and control groups. For researchers, such Randomised Control Trials (RCTs) are often considered the ‘gold standard’ of evaluation. Properly implemented, randomisation ensures that treatment and control groups are comparable both in terms of observed and unobserved attributes, thus identifying the causal impact of the project. However, implementation of these ‘real world’ experiments is challenging and can be problematic. RCTs may not always be feasible for local economic growth policies – for example, policy makers may understandably be unwilling to randomise the location of projects.19

Where randomised control trials are not an option, ‘quasi-experimental’ approaches of randomisation can help. These strategies can deal with selection on unobservables, by (say) exploiting institutional rules and processes that result in some locations quasi-randomly undertaking projects.

Even using these strategies, though, the treatment and control groups may not be fully comparable in terms of observables. Statistical techniques such as Ordinary Least Squares (OLS) and matching can be used to address this problem.

Note that higher quality impact evaluation first uses identification strategies to construct a control group and deal with selection on unobservables. Then it tries to control for remaining differences in observable characteristics. It is the combination that is particularly powerful: OLS or matching alone raise concerns about the extent to which unobservable characteristics determine both treatment and outcomes and thus bias the evaluation.

---

Evidence included in the review

We include any evaluation that compares outcomes for people or firms taking part in apprenticeship programmes (the treated group) after the programme with outcomes in the treated group before the programme; relative to a comparison group used to provide a counterfactual of what would have happened to these outcomes in the absence of the programme.

19 Gibbons, Nathan and Overman (2014).
This means we look at evaluations that do a reasonable job of estimating the impact of the project using either randomised control trials, quasi-random variation or statistical techniques (such as OLS and matching) that help make treatment and control groups comparable. We view these evaluations as providing credible impact evaluation in the sense that they identify effects that can be attributed, with a reasonable degree of certainty, to the project in question. A full list of shortlisted studies is given in Appendix A.

**Evidence excluded from the review**

We exclude evaluations that provide a simple before and after comparison only for those people or firms taking part in apprenticeship programmes, because we cannot be reasonably sure that changes for the treated group can be attributed to the effect of the project.

We also exclude case studies, cross-sectional participant surveys or evaluations that focus on process (how the project is implemented) rather than impact (what was the effect of the project). Such studies have useful roles to play in helping formulate better policy but they are not the focus of our evidence reviews.
To identify robust evaluation evidence on the causal impact of apprenticeships we conducted a systematic review of the evidence from the UK and across the world. Our review followed a five-stage process: scope, search, sift, score and synthesise.

Stage 1: Scope of Review

Working with our User Panel and a member of our Academic Panel, we agreed the review question, key terms and inclusion criteria. We also used existing literature reviews and meta-analyses to inform our thinking.
Stage 2: Searching for Evaluations

We searched for evaluation evidence across a wide range of sources, from peer-reviewed academic research to government evaluations and think tank reports. Specifically, we looked at academic databases (such as EconLit, Web of Science and Google Scholar), specialist research institutes (such as CEPR and IZA), UK central and local government departments, and work done by think tanks (such as the OECD, ILO, ippr and Policy Exchange.) We also issued a call for evidence via our mailing list and social media. This search found just over 1250 books, articles and reports (the full list of search terms can be found online here: whatworksgrowth.org/policies/apprenticeships/search-terms)

Stage 3: Sifting Evaluations

We screened our long-list on relevance, geography, language and methods, keeping impact evaluations from the UK and other OECD countries, with no time restrictions on when the evaluation was done. We focused on English-language studies, but would consider key evidence if it was in other languages. We then screened the remaining evaluations on the robustness of their research methods, keeping only the more robust impact evaluations. We used an adjusted version of the Maryland Scientific Methods Scale (SMS) to do this.\(^20\) The SMS is a five-point scale ranging from 1, for evaluations based on simple cross sectional correlations, to 5 for randomised control trials (see Box 2). We shortlisted all those impact evaluations that could potentially score three or above on the SMS\(^21\). In this case we found one evaluation scoring five: for examples of impact evaluations that score three or four on the SMS scale see the case studies and our scoring guide available at www.whatworksgrowth.org.

Stage 4: Scoring Evaluations

We conducted a full appraisal of each evaluation on the shortlist, collecting key results and using the SMS to give a final score for evaluations that reflected both the quality of methods chosen and quality of implementation (which can be lower than claimed by some authors). Scoring and shortlisting decisions were cross-checked with the academic panel member and the core team at LSE. The final list of included studies and their reference numbers (used in the rest of this report) can be found in Appendix A.

Stage 5: Synthesising Evaluations

We drew together our findings, combining material from our evaluations and the existing literature.


\(^{21}\) Sherman et al. (1998) also suggest that SMS Level 3 is the minimum level required for a reasonable accuracy of results.
Box 2: Our robustness scores (based on adjusted Maryland Scientific Methods Scale)

**Level 1:** Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. No use of control variables in statistical analysis to adjust for differences between treated and untreated groups or periods.

**Level 2:** Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group. In (a), control variables or matching techniques used to account for cross-sectional differences between treated and controls groups. In (b), control variables are used to account for before-and-after changes in macro level factors.

**Level 3:** Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g. difference in difference). Justification given to choice of comparator group that is argued to be similar to the treatment group. Evidence presented on comparability of treatment and control groups. Techniques such as regression and (propensity score) matching may be used to adjust for difference between treated and untreated groups, but there are likely to be important unobserved differences remaining.

**Level 4:** Quasi-randomness in treatment is exploited, so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment. This often entails the use of an instrument or discontinuity in treatment, the suitability of which should be adequately demonstrated and defended.

**Level 5:** Reserved for research designs that involve explicit randomisation into treatment and control groups, with Randomised Control Trials (RCTs) providing the definitive example. Extensive evidence provided on comparability of treatment and control groups, showing no significant differences in terms of levels or trends. Control variables may be used to adjust for treatment and control group differences, but this adjustment should not have a large impact on the main results. Attention paid to problems of selective attrition from randomly assigned groups, which is shown to be of negligible importance. There should be limited or, ideally, no occurrence of ‘contamination’ of the control group with the treatment.

Note: These levels are based on but not identical to the original Maryland SMS. The levels here are generally a little stricter than the original scale to help to clearly separate levels 3, 4 and 5 which form the basis for our evidence reviews.
Apprenticeships are not easy to define, and delivery models vary across countries, making definition harder. However, we can identify certain common features. For the purposes of this report, apprenticeships are defined as:

**Paid employment within a firm, alongside theoretical training that is usually provided by government, the employer, or a trade union, targeted specifically at school leavers (level of education varies by type of apprenticeship scheme). The apprentice often acquires a formal qualification by the end of the apprenticeship.**

Apprenticeships are not simply a) programmes which take place either entirely in the classroom (i.e. vocational education); or b) programmes that take place entirely in the firm (as discussed in the Employment Training Review). Rather, it is the combination of these two elements that matters. Although apprenticeship schemes do not necessarily need government support, such support is very common. In line with this, for all those programmes for which the evaluation provides details, there is always an element of public sector support.

**Impact evaluation for apprenticeships**

Evaluating the causal impacts of apprenticeship programmes is not straightforward. Ideally we would want to randomize participants into a programme, and then compare changes in their post-programme labour market outcomes with changes for similar people who did not participate. In principle randomisation is feasible for programmes like apprenticeships - however, we only found one example in this review (we found rather more in our review of other employment training programmes (T01).

Apprenticeships are, in many ways, a partnership between apprentice and employer, and unlike many employment training programmes, employers choose to get involved. As noted in the introduction, survey evidence suggests that participating firms are very positive about apprenticeships. This may reflect real benefits of such programmes, but also suggests that participating employers may not be representative of all employers (i.e., they ‘select into’ providing apprenticeships). In principle

---

22 These programmes are discussed in our employment training review: http://www.whatworksgrowth.org/policies/employment-training/.

23 BIS (2013).
randomising the firms who participate in apprenticeships is possible, but we found no real-world examples of programmes where this happens.

In the absence of randomization (of participants and/or firms), we worry about the ‘selection into treatment’ problem. As with active labour market programmes as a whole, selection into treatment occurs when individuals – or firms – participating in the programmed differ from those who do not participate in the programme, in ways that can be hard for researchers to observe.

For example, employers or programme delivery agencies may ‘cherry pick’ participants with the most skills or motivation to succeed in an apprenticeship. Agencies may also select the firms who provide the best opportunities. This means that the average effects of the apprenticeship are biased upwards, since the programme would not deliver the same benefits to other participants or in other businesses. Conversely, we may see downward bias, if programmes are targeted at the ‘hardest to help’ – such as people who are very long term unemployed or who have chaotic lifestyles, which make it harder for them to complete the apprenticeship.

A second set of challenges arises post-programme. A number of the studies here set out to explore the impact of ‘post-apprenticeship events’, such as moving jobs or career paths. Just as participants may select into an apprenticeship, however, their decision to stay or move jobs afterwards may be related to unobservable characteristics of those individuals. For instance, firms may actively seek to ‘lose’ the ‘worst’ apprentices.

A third set of challenges is more prosaic. Apprenticeships can have long timescales – sometimes running for three years – which means that some participants may drop out during the programme. Keeping track of participants post-programme raises similar issues. It is not clear how long any effects of apprenticeships might last – in our shortlist we have studies tracking immediate impacts (such as the first job gained), as well as lifecycle effects (wage gains over the following 40 years). Understanding the longer-term impacts of apprenticeships is crucial for policymakers – but is challenging to do in practice. Non-participants – in control groups – may be even harder to track, especially over long time periods.

The most robust studies in our shortlist adopt imaginative strategies to deal with these challenges, and to establish treatment and control settings. One paper (study 994) uses a Randomised Control Trial approach and scores 5 on the Maryland Scale. This paper looks at the US Community Restitution Apprenticeship Focused Training initiative (CRAFT), a 6-month employment programme designed to train and place high risk youths and juvenile offenders in employment in the construction industry.

Three papers score four on the Maryland Scale, and use a range of approaches for identification. One uses instrumental variables (IVs). Study 878 looks at wages, investigating the effect of the Austrian apprenticeship system, which is a one-year vocational training scheme aimed at secondary school leavers. It focuses on firms who go bankrupt after taking on apprentices, so that apprentices’ training time varies in ways participants cannot control. Study 897 uses spatial and time variation in policy rollout in Italy to identify effects: apprenticeships legislation was introduced in different regions at different times, unrelated to regions’ underlying economic trends. As a result, the evaluation is able to compare the change in performance of similar firms ‘treated’ with apprenticeships at different points in time. Study 939, another Italian evaluation, uses a regression discontinuity design that exploits different regional age cutoffs in eligibility for apprenticeships. As a result of these variations in eligibility similar people under the age of (say) 27 will get apprenticeships in some regions and not others.

Studies look at post-apprenticeship job moves tend to focus on particular classes of movers, where the decision to move is (more or less) random. For example, Study 867, for Germany, looks only at apprentices who move job when their employer shuts down or conducts a mass layoff. In some cases these movers also have to look for work in different industries to which they trained, which helps test the transferability of their skills and human capital.

These examples show a number of ideas that UK policymakers could adopt to evaluate the apprenticeship system in this country. In addition, access to administrative datasets that track individuals over time – such as data held by the HMRC, DWP and DfE – will be important in identifying the long term impacts of apprenticeship programmes.

The returns to education literature

The papers we review are those where apprenticeships are the main focus of the evaluation. This helps ensure that the studies we consider explicitly tackle the challenge of trying to identify the causal impact of apprenticeships on outcomes such as employment and wages.

The apprenticeships studies we review here complement the wider literature on the returns to education. The latter typically regresses wages on educational achievement controlling for personal characteristics (i.e. estimates a Mincerian wage equation). As with the studies we consider on apprenticeships, these studies struggle with selection bias: there is selection into both type and length of education based on unobservable individual characteristics such as motivation and ability. In order to deal with the selection problem, a large body of ‘returns’ studies use instrumental variable approaches. Card (1999) summarises some of these earlier IV studies, finding the wage return to a year of schooling to typically lie between 8 and 13%.

In contrast to the extensive general returns to education literature, there are far fewer studies where the main focus is apprenticeships. To the extent that the strategies used in the return to education studies convincingly deal with selection in to apprenticeships they would provide additional evidence on wage effects. But assessing this would be a time consuming task – hence our decision to focus on studies explicitly looking at apprenticeships and to cover a wider range of outcomes than in some of our other studies.

Most of the available evidence that we review uses a before-and-after comparison against a control group (SMS 3). A before-and-after study of the wage returns to apprenticeship is challenging for two reasons. First, it is difficult to find a pre-apprenticeship wage as most apprentices are school leavers and did not work before. Second, the before-and-after comparison should cover a long study period since apprenticeships are a long term investment in which the pay of materialise over the lifecycle of earnings.

For these reasons we only find relatively few studies (five in total) that estimate the wage returns to apprenticeships. Three of these use an instrumental variable approach, as in the wider returns to education literature (837, 878, 986). Since a pre-apprentice wage is not available, the two studies that use a before-and-after approach (847, 976) use wage changes at a later stage in the apprentice’s career. This is not a strict before-and-after comparison but does eliminate the effect of fixed unobservables. The drawback here is that it identifies only the effect on wage changes not levels; thus only a partial impact. These wage return studies typically study a longer period than evaluation of other outcomes. For example a dataset covering over 20 years is used in study 837.

Some studies examine a different aspect of wages by looking at the change in wages following a job

25 See Appendix B for the key words used to identify studies.
move for apprentices who move out of their profession against those who stay in the profession. This provides a measure of transferability of skills rather than general returns to apprenticeships.

Many of the studies we review, however, do not examine wages at all but instead consider outcomes such as employment and further study. These are more easily implemented as a before-and-after policy evaluation with shorter evaluation periods. More than perhaps any of our other reviews, the existence of a substantial evidence base on the returns to education suggests that the focus in this policy area needs to shift to assessing the effectiveness of different elements of policy design. We return to this issue below.
Findings

This section sets out the review’s findings. We begin with a discussion of the evidence base, and then explore the overall pattern of results. After this we consider specific outcomes in more detail.

Quantity and quality of the evidence base

The review initially considered over 1,250 articles and evaluations from the UK and other OECD countries, which were picked up under the initial keyword search. On further high level review, almost 1,000 were sifted out as not ultimately relevant (because they were theoretical rather than data-based, because they were comparative or descriptive rather than analytical, or because they reviewed non-OECD countries, were written in a foreign language or because of subject relevance for example). 74 were discarded as purely qualitative evaluations leaving some 171 articles were shortlisted for detailed review.

A further 61 of these 171 were ultimately discounted on grounds of relevance, and 83 were discounted on the grounds of not meeting the Centre’s minimum standard of evidence. The remaining 27 evaluations have been included in this review. Relative to our other reviews, this is roughly in the middle in terms of the quantity of impact evaluation evidence available. Table 1 shows the distribution of the studies ranked by SMS score.

<table>
<thead>
<tr>
<th>SMS Score</th>
<th>No. of studies</th>
<th>Evaluation reference numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMS 3</td>
<td>23</td>
<td>837, 838, 847, 859, 862, 867, 876, 880, 900, 904, 937, 942, 950, 960, 964, 976, 979, 985, 986, 987, 990, 997, 1045</td>
</tr>
<tr>
<td>SMS 4</td>
<td>3</td>
<td>878, 897, 939</td>
</tr>
<tr>
<td>SMS 5</td>
<td>1</td>
<td>994</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

At the time of writing areas covered and number of impact evaluations considered are as follows: employment training (71); sports and culture (36); road and rail (29); access to finance (27); business advice (23); estate renewal (21) and broadband (16). See http://www.whatworksgrowth.org/policies/ for an up-to-date comparison.
Country coverage

The papers evaluated in this report cover a number of different countries, mostly European. Table 2 provides a country breakdown.

Over a third of the papers investigate the German Apprenticeship System. A further seven papers examine apprenticeships in the UK and United States, with others looking at a variety of different European countries, plus one study for Australia.

As discussed in the introduction, variation by country complicates comparisons of the evaluation evidence. Where necessary, we break out results by broad type of scheme according to the three way classification – outlined in the introduction.

<table>
<thead>
<tr>
<th>Country</th>
<th>No. of evaluations</th>
<th>Study numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>1</td>
<td>878</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
<td>937</td>
</tr>
<tr>
<td>Denmark</td>
<td>2</td>
<td>942, 990</td>
</tr>
<tr>
<td>France</td>
<td>1</td>
<td>838</td>
</tr>
<tr>
<td>Germany</td>
<td>11</td>
<td>837, 847, 867, 876, 880, 904, 950, 979, 986, 987, 997</td>
</tr>
<tr>
<td>Hungary</td>
<td>1</td>
<td>960</td>
</tr>
<tr>
<td>Italy</td>
<td>2</td>
<td>897, 939</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
<td>1045</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>3</td>
<td>862, 900, 985</td>
</tr>
<tr>
<td>United States</td>
<td>4</td>
<td>859, 964, 976, 994</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Focus of the evaluations

We consider evaluations that cover three aspects of the impact of apprenticeships. Some studies look at the impact of apprenticeships on individual or firm-level economic outcomes – these have been the main focus in our other evidence reviews. A second set of studies explore the effect of individuals’ post-apprenticeship decisions on their labour market prospects. Finally, a third group look at the factors influencing uptake of apprenticeships.27

All three of types of evaluation provide useful information for policymakers – so we distinguish between them in what follows. Specifically, we classify evaluations into three types as follows:

- ‘Returns to apprenticeships’: *Studies that compare the impact of undertaking an apprenticeship vs not undertaking an apprenticeship on economic outcomes for individuals or firms.* This is the type of evaluation most commonly used in our reviews where changes in outcomes for a clear treatment group who undertook an apprenticeship are compared to a similar control group who did not.

---

27 This breadth of impact evaluation evidence is unusual in our reviews to date, and reflects the large and long-standing labour and education economics evaluation literature.
‘Post-apprenticeship’: Studies that examine the effects on apprentices of ‘post-apprenticeship’ events, e.g. differences in economic outcomes for stayers vs movers. In the vast majority of cases, these papers evaluate the effects of a change in employer or occupation post-apprenticeship, and evaluate effects on those programme participants that have moved from their training firm vs those that stay with their training firm.

‘Take-up’: Papers that evaluate the effects of specific policy interventions upon apprenticeships. These papers take apprenticeships as the dependent variable, and normally involve an examination of the effect of various policies on apprenticeship supply and demand for example.

As with other evidence reviews, in terms of establishing whether policy is effective, we are most interested in the first group of evaluations – looking at the returns to apprenticeships. We start with these studies, before reviewing the evidence on post-apprenticeship and take-up.

Returns to Apprenticeships

Nineteen evaluations consider the ‘returns to apprenticeships’ – i.e. look at the effect of apprenticeships on one or more economic outcome. Most of these consider the impact on individuals in terms of employment or wages. A smaller number consider the effect on individual skills or the effect on firms. Given the diversity of outcomes we think that the results by outcome are more informative than an overall judgement on whether apprenticeships ‘work’.28

A breakdown of the studies by outcome is provided in table A1 in the appendix. We have separated out the firm and individual level studies and grouped the later in to three broad categories - human capital accumulation, wages, and employment and labour market effects.

Individual outcomes: skills and further training

There is some evidence that apprenticeships improve skill levels, and stimulate further training / study.

Apprenticeships are designed to raise participants’ formal skills and human capital, it is hoped, stimulate further training or study. Only three studies consider these effects (studies 838, 964, 994). The two studies that look at formal training found some positive effects – one (study 838) on likelihood of undertaking further study post-apprenticeship, the other (study 994) for attendance and length of participation in formal training, but with no effect on completion. Study 964 also looks at the likelihood of undertaking further training post-apprenticeship, finding mixed results.

In terms of skill levels, study 994, which evaluates the US CRAFT programme described above, looks at formal training during the programme. Using a Randomised Control Trial, the study finds that CRAFT participants are significantly more likely to have attended a General Diploma programme. Apprentices attending the General Diploma programme also tend to participate for longer (i.e. if they drop-out, the drop-out later). However, high school graduation rates did not differ between treatment and control groups. Study 838, on the French apprenticeship system, finds that programme participants are significantly more likely to get a high school diploma, and also have a 42% higher probability of staying in education after finishing school.

28 In some of our other reviews, where there is less diversity in outcome, but more studies that look at multiple outcomes, we have found it useful to summarise the overall pattern of findings (to account for possible correlation across outcomes for evaluations of a specific programme that considers multiple related outcomes).
Paper 964 also looks at likelihood of undertaking post-apprenticeship training. It evaluates the US School to Work Programme, which is aimed at ‘hard to help’ groups. The authors find mixed results on the likelihood that participants partake in further training. For men who are likely to attend college there is a negative effect on two-year college attendance, while for men unlikely to attend there is a positive effect (although a negative effect on four-year college attendance). For women, there is a positive effect on both two-year and four-year college attendance. Unfortunately, the authors do not provide further analysis on what might be driving this result.

Even if an apprenticeship does not lead to formal qualifications or further training it is still possible that they improve worker productivity (indeed, this is a central aim of apprenticeships). In turn, this increased productivity could translate into higher wages. We consider wage effects next.

**Individual outcomes: wages**

Apprenticeships can increase wages, although in a couple of evaluations some effects are negative. Impacts also vary by type of participant.

If apprentices are more productive this could translate into higher wages, either at the firm which trains them or more broadly through their future careers. Offsetting this, apprentices’ wage bargaining power may be limited allowing firms, rather than workers to benefit from any productivity improvements. Both the extent of any productivity effects and variations in bargaining power may mean that wage effects vary across apprenticeships.

Five studies explore wage effects, with only one (study 837) finding consistently positive effects. Paper 837 finds that apprenticeships lead to wage gains in the 40 years after the programme. Study 878, using Austrian data, finds a wage gain of around 2.6% for a one-year apprenticeship scheme, and only for participants in firms with less than 10 employees; wage effects for apprentices in all other firms are statistically insignificant. Study 847 also finds mixed effects of apprenticeships upon wages, dependent on the schooling system undertaken prior to apprenticeship; for apprentices taking the Abitur route (pre-university route) the effects are negative, whilst for those taking the Hauptschule route (pre-apprenticeship route) the effects are positive.

Studies 976 and 986 find that apprenticeships have a statistically insignificant effect on participants’ wages. Study 986 also uses German data, and compares wages for 23-26 year olds on apprenticeships versus those on other vocational training schemes, finding no significant difference between the two groups. Study 976 finds no relationship between the duration of any form of apprenticeships and wage levels in the US.

29 The likelihood of attending college is estimated using a reduced form model for college attendance estimated excluding information on school-to-work participation. Individuals are ‘likely to attend’ if they are in the upper half of the distribution of predicted probabilities with respect to their gender.

30 The paper actually demonstrates net positive effects on a measure or ‘welfare’ (in the sense used by economists, rather than welfare payments). The German apprenticeship system, which lasts three years, pays participants notably lower wages compared to other countries, and these lifecycle gains in ‘welfare’ take into account the short term opportunity costs of participating in the programme. Given, however, that the after-programme gains largely reflect increases in wages, we include this study in this section.
Evidence Review: Apprenticeships - September 2015

Individual outcomes: employment and labour market

Apprenticeships tend to have a positive effect on participants’ subsequent employment (and also reduce unemployment post-programme).

In contrast to the effect on wages, the 11 evaluations that consider post-programme employment, generally find positive effects. Nine of these studies look at the effect of apprenticeship programmes on participants’ subsequent labour market outcomes. Five studies (900, 964, 965, 985, 994) look directly at employment after the apprenticeship is completed, with three (960, 985, 994) finding positive effects and two, finding more mixed results.

Another four studies look at unemployment post-apprenticeship (studies 862, 880, 986 and 1045); all four find that apprenticeships reduce the chances being unemployed.31

The final two studies consider specific aspects of employment. Study 939 focuses on apprentices’ subsequent entry into the job market, finding positive effects of programme participation on landing a first job which is a permanent position (either in the same firm or somewhere else).

Study 997 looks at career mobility post-apprenticeship, specifically apprenticeship participants’ movements from low-skilled to higher-skilled occupations. Using German data, it finds positive effects.

Firm level outcomes

There is some evidence that firms participating in apprenticeships experience economic gains, such as higher productivity or profits.

There are only two studies in the shortlist that look at the impacts of apprenticeships on participating firms.

Paper 897 investigates the impact of the Biagi Law (2003) which saw changes to the Italian apprenticeship system through raising the maximum age to 30 and introducing the option for workplace training. This evaluation finds positive effects of this legislative change on firm level employment, with the apprenticeship contract reform causing an increase in the level of apprenticeship employment of 5.2%.

Paper 987, on the other hand, which looks at the Südwestmetall employer association coaching programme in Germany that targets disadvantaged young adults and commissioned a training provider to administer the programme and match firms and disadvantaged youths, had no impact upon firm level employment.

These conclusions broadly match with evidence from surveys of participating employers, almost all of which suggest strong support for programmes (albeit on a selected sample of employers).32

Returns to apprenticeships: breakdown by country

Can these results tell us anything about whether effects differ across apprenticeship systems? To help answer this question, we use the information on Table 2 on the breakdown of studies by country. The evaluation of wage effects and of labour market effects are the only outcomes for which we have sufficient studies to contemplate differences across countries.

31 Study 862 looks at exits into a range of labour market states including employment and unemployment, but the focus of the paper is on the latter.
32 For a recent UK example, see BIS (2013). Lerman (2014) provides a review of the international survey evidence.
For wages, four out of the seven evaluations come from Germany. Two of these find unambiguously positive results, but one reports mixed results and one no effect. In contrast of the three remaining studies one shows no effect (for the US) and two show mixed effects (for Austria and Germany). Overall, given the small number of studies available and the split across countries it would be hard to conclude that there is strong evidence of marked differences in terms of the effect on wages.

Turning to other labour market effects, we can group together the nine studies looking at employment and unemployment. These evaluations have a more balanced distribution across countries with two covering Germany (both find positive effects), three covering the UK (two find positive effect, one finds no effect) and one each covering Hungary (positive), Norway (positive) and the US (no effect). Again, given the small number of studies available and the split across countries, it would be hard to conclude that there is strong evidence of marked differences in terms of the effect on employment.

While the distribution of evaluations across countries makes it hard to reach any firm conclusion on differences between systems, the geographical spread does give us some confidence that the findings are generalizable outside a single context.

Programme design

The country level comparisons that we have considered so far involve contrasting results from different systems. In this section, we take a different approach and consider whether there are any specific aspects of programme features that are correlated with programme success. It is important to recognize that the small amount of evidence restricts the extent to which we can control for other aspects of the system which may also affect outcomes. As before, given the number of studies available, we focus our discussion only on effects on wages and labour market outcomes (employment and unemployment).

Duration

The effect of the duration of the apprenticeship on wages or employment is unclear.

The reported lengths of each apprenticeship programme varied. In some cases, where no specific duration of apprenticeship was given, we assign programme duration based on country (e.g. we assume that German apprenticeships last 3-4 years).

As with country comparisons as a whole, there are no strong pattern of differences across apprenticeships of different duration. For wages, apprenticeships lasting 3-4 years show a mixture of positive (studies 837); mixed (studies 847 and 878) and no effects (study 986).  

For employment results are similarly inconclusive. Both studies that look at apprenticeships lasting 3-4 years show positive effects (studies 880 and 986). But so do two out of the three studies looking at apprenticeships lasting 2-3 years (studies 960 and 1045) and the one study that looks at apprenticeships of shorter than a year (study 994). Consistent with this, study 985 which considers multiple length apprenticeships also finds positive effects on employment.

This final evaluation (study 985) also provide a direct comparison of apprenticeships of different lengths – specifically the intermediate (NVQ Level 2) and advanced (NVQ Level 3) apprenticeships for the UK. The paper reports greater wage and employment returns for advanced apprenticeships,

33 Apprenticeship length could not be determined for the final evaluation that looks at wages (study 976)
34 For completeness note that the third study for apprenticeships lasting 2-3 years shows no effect (study 964), as does one study where apprenticeship length could not be determined (study 976).
which tend to have longer durations than intermediate programmes. The paper sets out the benefits to the individual and to the exchequer of both programmes. The Level 2 apprenticeship costs £5,202 to the exchequer, which is £2,013 less than Level 3 (which costs £7,215 unit cost). The benefits are seen to greatly outweigh these costs in both cases, with a return to the exchequer of between £31,000 and £48,000 at Level 2, and between £56,000 and £81,000 for Level 3 apprenticeships. At the individual level the lifetime benefits associated with the acquisition of Apprenticeships at Level 2 and 3 are very significant, standing at between £48,000 and £74,000 for Level 2 and between £77,000 and £117,000 for Level 3 Apprenticeships. It should be noted that this finding is only for one study and should be treated, therefore with the relevant caution.

**Different skill levels**

**On the (limited UK) evidence available, Level 3 and higher apprenticeships deliver substantially higher lifetime wage gains than lower level apprenticeships.**

It is important to note that findings on duration need to be interpreted with caution because longer programmes are likely to lead to higher skill levels and different formal qualifications. So duration may tell us something about ultimate qualifications, rather than simply the effect of programme length. Unfortunately, very few studies investigate skills levels of individual apprenticeship schemes and compare different qualifications obtained. However, one of the UK evaluations (study 985) does investigate the differences between different NVQ level qualifications.

Article 985, which looks at the impact of all British Apprenticeships upon earnings, finds that Level 3 apprentices and Level 2 apprentices see earning returns compared to similar individuals who don’t gain an apprenticeship of 22% and 12% respectively. It also finds that the effects on employment can differ depending on the type of apprenticeship. We reported on the relative cost-benefits above.

Article 900 also looks at skills levels, investigating the Youth Training Scheme which provides a range of training to unemployed 16-17 year olds. It looks at the impact of undertaking the YTS alongside an apprenticeships, versus undertaking YTS without an apprenticeship. The results suggest that there is no significant effect of any of these education programmes on unemployment.

Two other papers specify that a qualification is obtained as a result of their apprenticeship scheme: however they do not look at the impact of achieving different levels of qualification upon economic outcomes.\(^35\)

**Sectoral targeting and outcomes**

**There is too little evaluation evidence to draw clear conclusions on whether apprenticeships work better in some sectors than others.**

Very few papers provide details on the extent to which schemes target particular sectors. We have one evaluation (study 994) that looks at employment effects for a scheme aimed at the construction industry (it finds positive effects; also for course attendance). However, only one evaluation, for Hungary, looks at the return to apprenticeships in different sectors. It finds that the probability of being employed differs a lot between industries, but that the positive effect of workplace-based training compared to school-based vocational training is stable across industries.\(^36\)

\(^35\) Paper 838 for France; paper 937 for Australia.

\(^36\) Paper 904 also considers difference across manufacturing and service industry for post-apprenticeship moves – see below.
Apprenticeships versus other forms of training

There is some evidence that apprenticeships are more likely to increase employment than other forms of employment training (unless that training also involves an in-firm element). The evidence on wages is more mixed and appears to vary by gender.

Our employment training review found 71 evaluations that look at the returns to employment training. We have a considerably smaller evidence base for apprenticeships (19 evaluations). For apprenticeships, of the 9 evaluations that look at either employment or unemployment, 7 have positive effects. Taken at face value this is 77% of evaluations and compares favourably to the 60% of employment training studies that found positive effects on employment (40 out of 67). Interestingly, however, when we specifically focused on employment training that involves a within firm element we found only 4 studies (out of 17) reporting no or negative effects, while 13 reported positive findings. As in the employment training review, we think that the main message to emerge here concerns the importance of involving firms in the training process. In-firm / on the job training programmes outperform classroom-based training programmes. Employer co-design and activities that closely mirror actual jobs appear to be key design elements.

Three papers make specific comparisons between apprenticeships and other forms of employment training. Study 900 compares employment effects from UK apprenticeships, the Youth Training Scheme (YTS), apprenticeships and YTS together. The study, however, finds no significant effect on unemployment for any of these variants.

Study 964 looks at the US ‘School to Work’ Programme, which was aimed at disadvantaged young people and included a number of strands including job shadowing, mentoring, combined academic and vocational education; work placements, single career-focused study and apprenticeships. The study finds that for male participants, a number of strands are effective, including apprenticeships, which raise employment and decrease inactivity after leaving secondary school. For women, apprenticeships are the most effective strand in terms of raising wages.

Study 986 compares earnings and labour market participation for those taking German apprenticeships versus other forms of training in the country. Apprenticeship participation leads to substantially lower unemployment rates compared to other forms of vocational training, although this effect diminishes over time. Interestingly, the study finds no difference in wages between participants in apprenticeships versus those who did other types of vocational training. This is consistent with the bigger picture on apprenticeships and wages, discussed above, which finds mixed effects on post-programme participant earnings.

Post Apprenticeship decisions

There is some evidence that post-apprenticeship moves can increase wages although effects depend on circumstances (e.g. whether the individual chooses to move, stays in the same occupation, or moves within the manufacturing sector)

37 For sake of comparison we have treated employment training study 234 (classified as mixed) as positive: it finds positive effects on transition to employment post training, negative effects during training. Similarly for employment training study 236 (classified as mixed): it finds no effect on transition to employment during training, but positive effects for duration of future employment spells. Finally we treated employment training study 243 as zero or negative (classified as mixed) because effects on employment are negative for four years post training before reducing to zero.
This group of evaluations investigates the impacts of choices or events taken post-apprenticeship upon various economic indicators. For example the wage effect of an individual staying with the training firm post-apprenticeship vs. moving to a new firm. We found four studies of this kind – all for Germany – that met our minimum standards, and discuss each briefly below. In each case, the studies find ways to tackle the fact that moving / staying might be influenced by unobservable characteristics (such as ambition) or contextual factors (such as bargaining power or other conditions within a firm).

Study 876 looks at the factors influencing whether apprentices take their first job in the firm that trained them, or instead move to a different firm. Higher wages during the apprenticeship, higher training intensity and longer apprenticeship programmes increase the likelihood of staying with a firm post-programme. Stayers’ jobs tend to last longer than those of movers.

Three studies, all German, look at the wage effects of moving / staying. Study 950 provides the most detail on different types of moves. Overall, the wage penalty from moving job is close to zero, but this varies depending on why the move happens: for those who decide to quit there is a positive significant wage effect, but for those who lose their jobs there is no significant impact either way. Study 867 picks up on the latter issue and looks at wage outcomes for apprentices forced to move firm because of closure or layoffs, versus stayers. It finds that participants who stay within the same profession in their new job have higher wages, while those who shift trades end up with lower wages than before. Finally, study 904, also for Germany, finds some sectoral differences in wage outcomes; apprentices in manufacturing who move job experience wage gains, while those working in construction, crafts, commerce or trading occupations experience wage losses when moving jobs.

**Take-up of Apprenticeships**

There is some evidence that identifies mechanisms that may increase entry into apprenticeships and attendance during the programme (e.g. pre-qualifications, higher wages and subsidies to individuals). However, we have less evidence on what works to ensure people complete apprenticeships.

This group of papers, rather than looking at the impacts of apprenticeships, investigates the impact of various policies upon apprenticeship entry and completion.

Two studies look at tools that aim to raise entry into apprenticeships. Study 937, on Australia, finds that Certificate I/II pre-qualifications introduced under the country’s national qualifications increased the likelihood of participating in an apprenticeship or traineeship, both for men and women. Study 942, on a Danish pre-apprenticeship programme, found a positive and significant impact of the programme on participants entering an apprenticeship afterward.

Two studies look at factors affecting attendance during an apprenticeship programme. Study 876 looks at the level of wages paid to German apprentices, as well as the length of the apprenticeship; higher wages and longer programmes both have positive, significant effects on apprentice retention. Study 990 evaluated the Danish Apprenticeship Subsidy scheme (AAS), finding a significant positive effect on the vocational attendance rate among low-schooled 25-year-old men in its first full year of operation but no significant effect thereafter.

Three studies explore drop-out from programmes. Study 859 compares ‘regular’ apprenticeships in the US with ‘joint’ apprenticeships which are co-sponsored by the firm and a union. It finds mixed
results: those on joint programmes take longer to complete an apprenticeship (in terms of average months taken) but they are also less likely to drop out and have higher completion rates.

Study 979, for Germany, evaluates an apprenticeship bonus for employers, which rewards firms for taking on participants and when they complete the course – it finds no effect of this on participation or completion rates. This stands in contrast to study 876, which suggests payments to individuals (rather than firms) can be effective.

Finally study 838, for France, compares the rates of school dropout for apprentices versus regular students, finding no statistically significant difference between the two.
Summary of findings

This review considers the impacts of apprenticeships on workers and firms. This section summarises the detailed findings. We emphasise that many of these findings depend on a small number of studies. They are, however, consistent with other research on apprenticeships and on employment training more broadly (see our first evidence review).

What the evidence shows

• There is some evidence that apprenticeships improve skill levels, and stimulate further training / study.

• Apprenticeships can increase wages, although in a couple of evaluations effects are negative. Impacts also vary by type of participant.

• Apprenticeships tend to have a positive effect on participants’ subsequent employment (and also reduce unemployment post-programme).

• Higher level apprenticeships (Level 3 and above) deliver substantially higher lifetime wage gains relative to lower level apprenticeships (based on the limited UK evidence available).

• There is some evidence that apprenticeships are more likely to increase employment than other forms of employment training (unless that training also involves an in-firm element). The evidence of impact on wages is more mixed and appears to vary by gender.

• There is some evidence that identifies mechanisms that may increase entry into apprenticeships and attendance during the programme (e.g. pre-qualifications, higher wages and subsidies to individuals). However, we have less evidence on what works to ensure people complete apprenticeships.

Where the evidence is unclear

• It is unclear whether the duration of the apprenticeship matters for effects on wages or employment (although longer apprenticeships that deliver higher qualifications may have more positive effects)
Where there is a lack of evidence

- There is some limited evidence that firms participating in apprenticeships experience economic gains, such as higher productivity or profits.
- There is too little evaluation evidence to draw clear conclusions on whether apprenticeships work better in some sectors than others.
- There is some evidence that post-apprenticeship moves can increase wages although effects depend on circumstances (e.g. whether the individual chooses to move, stays in the same occupation, or moves within the manufacturing sector). However, these results are each based on a single study.
- There is no impact evaluation evidence looking at the effect of apprenticeships on a given local area (rather than individual participants or firms).
- There is no impact evaluation evidence comparing the effects of nationally run programmes versus locally run programmes.
- Existing ex ante modelling suggests that the economic benefits of apprenticeships comfortably outweigh their costs. However, only one of the impact evaluations provides cost data in a form which allows us to calculate ex-post benefit-cost ratios for that programme.
- None of the shortlisted studies look at the effects of substantially scaling up apprenticeship provision, as is currently happening in the UK. We need more evidence on whether identified benefits also hold in a larger programme. Given the other substantial changes to the UK apprenticeship system in the past decade and a half, more up to date UK impact evaluation evidence is also needed.

How to use these reviews

- Apprenticeships are currently very high on the policy agenda, and the evidence review highlights a number of factors for policy makers to be aware of when considering apprenticeships:
  - While the evidence suggests that higher level apprenticeships (level 3 and above) may offer better outcomes, it does not currently tell us whether this is because stronger candidates gravitate towards more demanding programmes. If this is the case, policymakers need to consider how to address the needs of those ‘left behind’ by this type of apprenticeship offering.
  - Any policy should carefully consider how to recruit firms to provide apprenticeships, and trainees to fill them. A better understanding of the costs and benefits to firms will help in this (see below). As will a better understanding of which policy design aspects increase take-up and reduce drop-out.

Helping to fill the evidence gaps

As should be clear from this review, there are many things that we do not know about the local economic impact of apprenticeships.

**If achieving local economic impact is an important part of the case for apprenticeship provision, then there need to be more evaluations that explicitly explore these impacts and how to maximize them.** Central and local policymakers – and private sector partners – should:
• Look to undertake systematic comparisons that cover different kinds of apprenticeship model - for example, the German system versus a more decentralised system.

• Conduct further research looking at outcomes for firms. Surveys of firms who offer apprenticeships suggest those firms see clear benefits, but they may not be representative of all employers.

• Set up evaluations of scheme design and its effect on take-up, completion and outcomes. This is particularly important given devolution of skills budgets to cities such as London and Manchester. Central and local policymakers should work together to design robust evaluation that increases our understanding of how to improve the design of apprenticeships.

• Make scheme cost data available to researchers so that robust benefit-cost ratios can be calculated.

The Centre’s longer term objectives are to ensure that robust evidence is embedded in the development of policy, that these polices are effectively evaluated and that feedback is used to improve them. To achieve these objectives we want to:

• Work with local decision makers to improve evaluation standards so that we can learn more about what policies work, where.

• Set up a series of ‘demonstration projects’ to show how effective evaluation can work in practice.

Interested policymakers please get in touch.
References


Oliver, D (2012). Lower Level Qualifications as a Stepping Stone for Young People, National Centre for Vocational Education Research Occasional Paper, Adelaide: NCVER.
Appendix A: Findings by outcome

Table A1: Findings by outcome ‘returns to apprenticeship’

<table>
<thead>
<tr>
<th>Outcome Evaluated</th>
<th>Number</th>
<th>Evaluation reference numbers</th>
<th>Positive</th>
<th>No effect</th>
<th>Mixed</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>838, 964, 994</td>
<td>838, 994</td>
<td>964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course attendance</td>
<td>1</td>
<td>994</td>
<td>994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course completion</td>
<td>1</td>
<td>838</td>
<td>838</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further education</td>
<td>1</td>
<td>838</td>
<td>838</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further Training</td>
<td>1</td>
<td>964</td>
<td></td>
<td></td>
<td></td>
<td>964</td>
</tr>
<tr>
<td><strong>Labour market outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moving into employment</td>
<td>5</td>
<td>900, 960, 964, 985, 994</td>
<td>960, 985, 994</td>
<td>900, 964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoiding unemployment</td>
<td>4</td>
<td>862, 880, 986, 1045</td>
<td>862, 880, 986, 1045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moving into permanent work</td>
<td>1</td>
<td>939</td>
<td>939</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moving to a qualified job</td>
<td>1</td>
<td>997</td>
<td></td>
<td></td>
<td></td>
<td>997</td>
</tr>
<tr>
<td><strong>Wages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>837, 847, 878, 976, 986</td>
<td>837</td>
<td>976, 986</td>
<td>847, 878,</td>
<td></td>
</tr>
<tr>
<td><strong>Firm Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm Level outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>897, 897, 987</td>
<td>897, 897</td>
<td>987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment (firm)</td>
<td>2</td>
<td>897, 897</td>
<td>897</td>
<td>987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td>1</td>
<td>897</td>
<td>897</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B: Evidence Reviewed

<table>
<thead>
<tr>
<th>Ref no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>937</td>
<td>Oliver, D (2012). Lower Level Qualifications as a Stepping Stone for Young People, National Centre for Vocational Education Research Occasional Paper, Adelaide: NCVER.</td>
</tr>
<tr>
<td>987</td>
<td>Mohrenweiser, J; Pfeiffer, F (2009). Coaching disadvantaged young people: Evidence from firm level data, ZEW discussion paper Number 14-054, pp27</td>
</tr>
<tr>
<td>994</td>
<td>Schaeffer, CM; Henggeler, SW; Ford, JD; Mann, M; Chang, R; Chapman, JE (2014). RCT of a promising vocational/employment program for high-risk juvenile offenders, Journal of Sunstance Abuse Treatment, Vol46, p114-148</td>
</tr>
<tr>
<td>1045</td>
<td>Storen, LI (2011). Key factors behind labour market marginalization of young immigrants: limited access to apprenticeships, ‘state dependence’ or low qualifications? Young, vol 19(2), p129-158</td>
</tr>
</tbody>
</table>

Find the full list of search terms we used to search for evaluations on our website here: whatworksgrowth.org/policies/apprenticeships/search-terms.
The What Works Centre for Local Economic Growth is a collaboration between the London School of Economics and Political Science (LSE), Centre for Cities and Arup.

www.whatworksgrowth.org