

WestminsterResearch

http://www.westminster.ac.uk/westminsterresearch

Evaluation of Education and Training Impacts for the Unemployed: Challenges of New Data Urwin, P. and Cergua, A.

NOTICE: this is the authors' version of a work that was accepted for publication in Labour Economics. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Labour Economics, DOI: 10.1016/j.labeco.2020.101907, 2020.

The final definitive version in Labour Economics is available online at:

https://doi.org/10.1016/j.labeco.2020.101907

© 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <u>https://creativecommons.org/licenses/by-nc-nd/4.0/</u>

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Evaluation of Education and Training Impacts for the Unemployed: Challenges of New Data

Third Revised Version

Augusto Cerqua^a, Peter Urwin^b, Dave Thomson^c and David Bibby^c

^a Department of Social Sciences and Economics, Sapienza University of Rome, P.le Aldo Moro, 5 - 00185 Rome, Italy

^b School of Organisations, Economy & Society, University of Westminster, 35 Marylebone Road, NW1 5LS, London, UK

 $^{\rm c}$ FFT Education Datalab, 11 Tufton Street, London SW1P 3QB, UK

1. Introduction

Recent decades have seen the active labour market programme (ALMP) evaluation literature expand substantially (for a recent review see Filges et al., 2016). The increased availability of rich datasets has allowed studies to utilise extensive time periods, both before and after treatment (Sianesi, 2004; Lechner et al., 2007; Biewen et al., 2014); there has been some improved ability to identify detail of the training undertaken (Frölich, 2004; Sianesi, 2008; Lechner et al., 2011); and studies have been able to give more explicit consideration to the issue of 'dropouts' in experimental (Heckman et al., 1998a) and non-experimental settings (Paul, 2015; Choe et al., 2015).

This has led to an increase in the variety of methods used by researchers; characteristics of the programmes studied, and richness, or otherwise, of data used to capture differences between programmes (Biewen et al., 2014). Drawing on findings from across studies, there is some consensus that in the short run, programmes exhibit negative employment impacts due to lock-in effects; the duration and size of any such lock-in effects vary across programmes, but generally it takes two to three years for 'meaningful' impacts to arise; and the extent to which training is mandated within a programme seems negatively related to estimated impacts (see for instance, Card et al., 2018). These studies also identify some consistent messaging on effect heterogeneity, with on-the-job training seen as more effective than classroom-based training.

Despite these important insights, many questions remain. McCall et al. (2016) note the lack of explicit consideration across studies of what constitutes training; they argue that researchers should focus on estimating separate impacts for different training types, in contexts where programmes cannot be 'killed' for political reasons; and, noting 'the strong evidence of heterogeneity in treatment effects' (page 480), underline the need for insights that inform the matching of subgroups of unemployed to different interventions. A strong argument is made for future research that identifies effect heterogeneity below programme level. As more data becomes available to researchers, there are greater opportunities for such insights.

However, as this paper shows, richer data brings its own challenges. We analyse information from an exceptionally rich dataset, containing detailed information on training interventions, allowing us to contribute to the development of this evidence base. More

specifically, the administrative data used in this study contain information for an English population of 2.3 million individuals with an unemployment benefits claim start date between 6th April 2006 and 5th April 2008. This has been linked to detailed information on all registered learning aims at English further education (FE) institutions between the 2002/2003 and 2012/2013 academic years. English FE Institutions are broadly equivalent to US Community Colleges,¹ and most training for the unemployed is undertaken in these FE Institutions.² We focus on training for the unemployed undertaken during the first 18 months of an unemployment spell, when training treatments are predominantly voluntary in nature: estimating effects for three categories of training, and focusing analysis on unemployed individuals aged between 25 and 55.

Relatively large numbers in both treatment and comparison groups allow us to use Coarsened Exact Matching (CEM) (see Iacus et al., 2011), to create month-by-month estimates of employment effects, up to 5 years on from training start. We match [via CEM] on a variety of characteristics, including socio-demographic variables, prior qualifications³, a flag of basic skill need from a caseworker and up to 8 years of labour market and learning histories. Studies suggest that matching on such histories, better ensures the Conditional Independence Assumption (CIA) is met (Heckman and Smith, 1999; Lechner and Wunsch, 2013; and Caliendo et al. 2017), when considering selection into treatment start. The fact that we observe all training spells undertaken in FE during the period under analysis, allows us to counter many of the concerns around substitution (see Fay, 1996; Heckman et al., 1998a; Heckman et al., 2000).

Our focus is on training interventions taken up voluntarily that can occur at different points in each individual's unemployment spell. This potential for variation across individuals in the timing of treatment, raises similar concerns to those flagged in studies by Sianesi (2004; 2008), Fitzenberger et al. (2006), Fredriksson and Johansson (2008) and Biewen et al. (2014) concerning dynamic selection into treatment. Therefore, in each of our analyses we separately estimate employment impacts for (a) those who initiate training in the first two months of their unemployment spell; (b) those who first initiate training in the third or fourth months of unemployment; (c) those who first initiate training for the first time in the seventeenth and eighteenth months from claim start. In the following analysis, when comparing outcomes for these individuals with those of the 'untreated', the latter group is made up of individuals who have not initiated training up to the specified point in time from claim start, but who may do so in future months. From here we therefore refer to this group

¹ The majority of learning undertaken by unemployed individuals in English FE Institutions is at, or below, 'Level 2' - as defined by the UK's National Qualification Framework (NQF). NQF Level 2 is equivalent to European Qualification Framework (EQF) *Level 3* or a US High School Diploma; and UK NQF Level 1 learning is equivalent to EQF *Level 2*. English FE has a focus on technical (vocational) programmes, similar to German Vocational Qualification Certificates and US Community College Certificates (and these courses are grouped as 'Full level 2 and above' in Table 1).

² UK government spending on FE has amounted to approximately £4bn per annum in recent years and the number of adult learners (aged 19+) participating in government-funded FE was 3.28m in 2012/13.

³ Combining a specific indicator within the FE administrative dataset that records prior qualifications (including those taken outside FE), with indicators obtained from a trawl of this administrative dataset to control for historical FE learning interventions.

as the *currently untreated*. Following Sianesi (2008), we present figures that are aggregates of these treatment effects estimated by time of initiation, weighted according to the observed distribution of initiation across the nine separate periods.

This approach allows us to tackle concerns over selection into treatment start. However, the more substantive contribution of this paper arises from (i) our ability to identify amongst the treated those who do, and those who do not, achieve the learning outcomes of training (according to the trainer's assessment, which may involve some form of examination); and (ii) our investigation of the challenge of selection into full and partial treatment, following training initiation, that incorporation of this information presents. As richer sub-programme data become available, it is possible to investigate the challenges posed by McCall et al. (2016), but identification of the drivers of effect heterogeneity below programme level brings us to a perspective more often associated with the Education Economics literature. Our investigation highlights one of the key points of possible convergence between an ALMP evaluation literature, where intention to treat approaches dominate; and an Education Economics literature, where achievers are often the central focus of attention in terms of estimated impacts⁴.

To investigate, we present estimated month-by-month employment effects following treatment start, using a variety of pair-wise comparisons; for each of our three categories of learning. First, we present the estimated employment impacts that arise from (A) comparison of all those who start a training treatment at a point in time [*all starters*], and those whom we do not observe initiating a treatment [*currently untreated*] up to that specified point in time, but who may do so in the future, up to 18 months from claim start date. Matching on lengthy employment histories and adopting the approach of Sianesi (2008), we invoke the Conditional Independence Assumption [CIA]. When comparing *all starters* with the *currently untreated*, the main concern is selection into treatment start, and in this respect the existing literature provides support for this invocation of the CIA (Caliendo et al., 2017).

However, we then present estimated employment impacts that arise from pairwise comparisons which utilise the information we have on achievement (or otherwise) of learning outcomes (i.e. splitting the group of *all starters* into *fully* and *partially* treated). Specifically, we present estimated impacts associated with comparisons between (B) the *fully treated* (who are recorded as achieving the learning aims of a training scheme) and the *currently untreated*; and (C) the *partially treated* (who enrol for this training, but are not recorded as achieving the learning outcomes) and the *currently untreated*. Estimands from analysis under (B) and (C) are obtained using a similar approach to those under (A), as the causal contrast is between 'training now' (whether this is for *all starters*, the *fully treated* or *partially treated* at a point in time) versus 'waiting' (i.e. those who at the same point in time are 'currently untreated').

Finally, we present estimated impacts from comparison between (D) the *fully treated* and the *partially treated*. Estimands obtained from this comparison differ to those under (A), (B) and (C) as unemployed individuals who start training during a given sub-interval and achieve

⁴ For a discussion of the former approach in the ALMP literature see for instance Heckman et al. (2000) and the approach that dominates in the Education Economics literature is epitomised by Blundell et al. (2005).

[*fully treated*], are compared to the unemployed who start training during the same subinterval, but do not subsequently achieve the learning outcomes [*partially treated*]. As a result, estimates obtained from (D) are closer to those used in standard cost-benefit frameworks, whilst this is not true of those under (A), (B) and (C).

McCall et al. (2016) suggest that these estimated impacts, 'represent the impact of completing the training and, hopefully, now possessing the set of skills the training aims to provide'. Completion is a proxy for the securing of human capital, and it would seem important to attempt to utilise indicators of achievement where we have them. Comparisons (B), (C) and (D) allow us to utilise the information on achievement, but also introduce the potential for bias associated with selection following training start.

It is likely that selection into full treatment and partial treatment is non-random, and there are differences between the two groups that are correlated with outcomes. Because selection into full or partial treatment occurs after treatment start, our identification strategy will only remain valid under certain assumptions. Namely, that matching up to the point of training start captures all relevant information driving subsequent selection into partial or full treatment. For this to be valid, any unobserved factors driving selection into these two groups, must be uncorrelated with potential outcomes, having conditioned on observed learning and employment histories prior to treatment start. However, if there are unobserved factors that drive selection into partial/full treatment, and this information only arises following treatment start, we are at risk of bias in our comparison of effects from *full* and *partial* treatment. Section 5.1 shows how our findings change when using distance to nearest FE training provider to instrument for selection into partial and full treatment – investigating the validity of estimates obtained under (B) to (D).

To ensure our distinction between partial and full treatment is located within the existing literatures, Section 2.1 considers the partial treatment framework of Heckman et al. (1998a); the literature on dropouts in ALMP studies (Kluve et al., 2012; Choe et al., 2015; Paul, 2015) and links this to the education economics literature that estimates impacts from vocational learning (Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Jepsen et al., 2014; Bibby et al. 2014; Hedges et al., 2018). Section 2.2 provides additional context, detailing the New Deal programme that was in operation during the period of this study. Section 3 describes the data and methodological approach, while Section 4 presents the results. Section 5 checks for the robustness and sensitivity of the estimates with a particular focus on the issue of dynamic selection following programme start (Section 5.1) and Section 6 summarises findings and implications for future research.

2. Theoretical Background and Labour Market Context

2.1 Framework for Consideration of Partial Treatment

A key focus of our study is the estimation of impacts arising from partial treatment and in the existing ALMP literature partial treatment is often synonymous with 'dropout'. Paul (2015) discusses all the possible implications of dropping out of a programme and shows how in theory this might lead to negative, as well as positive, effects on employment prospects with respect to fully treated individuals. One could imagine that less employable individuals

remain on a course for longer because they have fewer job offers. However, findings from studies of dropouts identify completers as having some of the highest impacts (likely from a combination of human capital impacts, and the signal gained from accreditation for a proportion of completers); whilst dropouts secure lesser, but still statistically significant, effects, likely from partial human capital enhancements (Heckman et al., 2000; Bibby et al., 2014; Hedges et al., 2018).

For instance, Choe et al. (2015) adopt the generalized propensity score to capture the extent of any effect from partial treatment for those who drop out. The study estimates a doseresponse function (DRF) of the length of training on the probability of employment, finding that dropouts benefit from longer training spells compared to shorter ones, but only after completing a certain duration threshold. Due to data limitations no consideration is given to non-participants, and the key identifying assumption is that selection into different lengths of training is based on a rich set of observed covariates, and any unobserved factors driving selection are uncorrelated with potential outcomes having conditioned on these observables. Fitzenberger et al. (2015) account for endogenous dropout, employing a flexible bivariate random effects probit model for employment and training status, estimated with Bayesian Markov Chain Monte Carlo (MCMC) techniques. This approach accounts for the potential endogeneity of program incidence and duration; and considers potential confounding from time-invariant unobservables and time-varying observed covariates. The author's findings imply positive effects of training on the employment probability for dropouts, emerging nine to twelve months after programme start in all subsamples considered.

However, an observed training duration that is less than that required to complete a course is an imperfect proxy for 'partial' treatment. There is always the possibility that a group of learners will contain individuals who do not formally drop out, but who do not engage in a way that ensures they are 'fully' treated. They may attend enough sessions to avoid being counted as a formal 'dropout', but these 'completers' are not fully treated. This may happen because of circumstances beyond their control (travel, personal circumstances etc.) or because the ALMP context makes 'observed' dropout undesirable (see, for instance, Arni et al., 2013; Boockmann et al., 2014), due to the financial incentives facing training providers and/or sanction regimes for the unemployed.

Consider for instance, Heckman et al. (2000) who suggest that, "unlike researchers conducting experiments in chemistry or biology, researchers conducting a social experiment have only partial control over the level of the treatment actually received by treatment and comparison group members". In the case of ALMP training interventions, if an individual does not formally 'drop out', it is not necessarily the case that they are engaged and fully treated. We may have many individuals who attend infrequently, but not infrequently enough to be formally recorded as dropouts and/or simply do not engage with the learning (i.e. they are 'unobserved dropouts')⁵.

⁵ Our approach is different to that suggested by [for instance] Behrman et al. (2004), who set out a more traditional notion of partial treatment which sees participants obtaining different quantities of treatment, and researchers attempting to estimate dose-response functions.

Heckman et al. (1998a) consider the challenges raised by the issue of partial treatment/dropout in the context of social experiments; proposing a simple representation, with Y_d capturing the potential impact of 'partial' treatment (from observed dropouts); and a potential full treatment impact Y_p , that can be estimated for those who are not observed to drop out of training. The authors focus discussion on the challenges in social experiments, where the presence of dropouts does not allow evaluation of the average impact of treatment, but only the average effect of the assignment to treatment, i.e. intention to treat. In this study of a non-experimental setting, we focus on the case where partial treatment for those who start the programme, can be due to either observed or 'unobserved' dropout.

The literature suggests that the percentage of individuals [observably] dropping out of training programmes, can range from 5% to 80% (Heckman et al., 2000; De Crombrugghe et al., 2010; Hirshleifer et al., 2015). However, across both social experiments and econometric evaluations, it is likely there remains substantial unobserved variation in the extent of partial engagement in training, as the level of observed dropout is an imperfect proxy – driven by rules governing the particular programme within which any training is located. In our study, the ability to distinguish unemployed individuals who do and do not achieve the learning outcomes of a programme, is equivalent to a distinction between *fully treated* (achievers), and *partially treated* (non-achievers) who may complete (unobserved dropouts) or not (observed dropouts). This discussion locates our study within the ALMP literature (for instance, Dorsett, 2006; Card et al., 2010; Kluve, 2010; Biewen et al., 2014), but we are also able to comment on a growing literature that estimates the impacts of vocational learning using the partially treated to create counterfactual outcomes (Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Jepsen et al., 2014; Bibby et al. 2014; Hedges et al., 2018).

Across these literatures concern centres on the extent to which comparison of outcomes for the fully and partially treated is confounded by non-random selection into these states, on factors that remain unobserved. For instance, Kluve et al. (2012) consider the extent to which duration of training is endogenously determined and find evidence to suggest that this is 'only relevant at the lower and upper part of the treatment duration distribution'. In Section 5.1 we investigate the issue of selection following programme start, using distance to nearest FE training provider to instrument for possible endogenous selection into partial and full treatment.

2.2 The UK New Deal

Since the 1980s, claimants of UK unemployment benefits have been required to actively seek work as a condition for benefit receipt (OECD, 2014). The eligibility conditions were further tightened with the introduction of the Jobseeker's Allowance (JSA)⁶ regime in 1996 and the New Deal in 1998. The New Deal was an ALMP introduced in the UK by the Labour government of Tony Blair, to provide training, subsidized employment and voluntary work for the unemployed; renamed as the Flexible New Deal (FND) between October 2009 to

⁶ JSA is a policy operated by multiple parts of the system, including the executive agency Jobcentre Plus, with bureaucratic and ministerial oversight provided by the Department for Work and Pensions.

June 2011, after which the current Work Programme was introduced.⁷ Individuals became eligible for the interventions provided under the New Deal following a certain duration of unemployment, which varied by specific target groups. For instance, those aged 18 to 24 were referred to the New Deal for Young People (NDYP) after 6 months of unemployment; whilst those aged 25+ were referred to the New Deal 25+ (ND25+) after 18 months of unemployment. Participation in the New Deal programmes began with an intensive period of job-search assistance (the Gateway phase), followed by the offer of training or alternative programmes (the New Deal options phase).

This paper focuses analysis on training and education undertaken in the period prior to any New Deal referral, when the take-up of training is voluntary in nature. We focus the analysis on individuals aged 25 to 55, during the 18-month period between their claim start date and mandatory ND25+ referral. We have also studied those aged 18 to 24, during the 6-month period prior to mandatory referral to the NDYP. However, the lack of rich employment histories for this younger age group hampers estimation of impacts that can be credibly interpreted as causal, and results are therefore relegated to the Appendix.

Our focus on the period prior to ND referral suggests we are capturing training events that may arise from discussion with Jobcentre Plus advisors, but during this period there are no sanctions for non-attendance. This locates our study closer to the contexts considered by US studies such as those by Heckman et al. (2000); Plesca and Smith, (2007); and Barnow and Smith (2015). It also enables us to relate the study to a branch of the education economics literature, that is predominantly focused on vocationally-oriented and Community College-based learning (Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Jepsen et al., 2014; Bibby et al. 2014; Hedges et al., 2018), undertaken in voluntary contexts. It also shares characteristics of the contexts studied by Sianesi (2004; 2008), as individual uptake of training can occur at any point during the 18-month period prior to ND referral.

To provide a wider context, we would refer readers to Bibby et al., (2015a; 2015b), where estimated impacts are presented for the period of New Deal following the period analysed here. These studies build on a literature that has developed since the introduction of the New Deal. For instance, Blundell et al. (2004) exploit area-based piloting and age-related eligibility rules to evaluate the effectiveness of the New Deal Gateway phase; and find that the programme raised transitions to employment by about five percentage points in the short run. Alternatively, Dorsett (2006) uses a propensity score matching approach to investigate the New Deal options phase, finding that a period of subsidized employment is a more effective means of exiting unemployment. Adopting a similar methodology, Dolton and Smith (2011) focus their attention on the effectiveness of a related but different unemployment policy in the UK - the New Deal for Lone Parents (NDLP), which was a large voluntary programme for single parents. Their estimates show large, and fairly persistent, effects of NDLP participation on the probability of benefit receipt. Lastly, Van den Berg et al. (2014) find that anticipatory effects on the hazard rate are present in the weeks before the onset of

⁷ The New Deal programmes were a major driver in the development of quasi-market structures in British employment services (OECD, 2014) due to the increasing delivery of modules by subcontracted providers.

ND treatment and in those weeks, individuals reduce their search effort. The potential for anticipatory effects of the New Deal are considered in Section 5.2.

3. Data and method

3.1 Data and Sample Selection

This study focuses on the population of unemployed individuals with a first "active benefit" claim start date between 6th April 2006 and 5th April 2008, who are resident in England and who may attend an FE institution in England to undertake training. We evaluate the impact of training taken up voluntarily, earlier in an unemployment spell prior to mandatory New Deal referral – at which point any observed training interventions become mandated. The focus of analysis is on the unemployed aged 25 to 55 during the period prior to the New Deal referral, which for this age group occurs when an unemployment spell has lasted for 18 months. We have information on all 2.3 million individuals flowing on to benefits between April 2006 and 2008, including information on up to eight years of prior labour market history⁸ and five years of outcomes.

The cohort of unemployed individuals is identified using information contained within the National Benefits Database (NBD). This data is then matched to information held on the Labour Market System (LMS), which is used by advisors within the English public employment service (Jobcentre Plus) and contains information on basic skill needs; the New Deal evaluation datasets that hold information on ALMP interventions; and administrative data from the Individualized Learner Record (ILR) which contains detailed information on all registered learning aims at English Further Education (FE) institutions between the 2002/2003 and 2012/2013 academic years. This information is then matched to administrative information from employment records (the Work and Pensions Longitudinal Study, WPLS). For each individual, the initial claim start date is considered as time (t) equal to zero, and their expected date of referral to ND is calculated from this. We trawl the LMS, ILR and New Deal datasets for all interventions/referrals (training or otherwise) that occur between time zero and the expected date of New Deal referral⁹.

Individuals starting a programme before New Deal referral mostly do so either because they are referred by caseworkers¹⁰, or because the individual self-selects into a programme. The LMS dataset contains valuable information on whether an individual has been identified as requiring some form of basic [skills] support and guidance, and this distinction is used in matching. Unemployed individuals are able to enrol on a large variety of different training schemes and classroom-based interventions. Using ILR data we classify these into three categories, which take into account the different levels of qualification:

⁸ As already suggested, we also present findings for analysis of unemployed individuals aged 18 to 24 but the lack of employment histories hampers estimation of impacts that can be credibly interpreted as causal.

⁹ In recognition of the potential margin for error around the expected claim start date [X], a 'fuzzy' X is created covering the period between X - 2 weeks and X + 2 weeks.

¹⁰ Caseworkers can encourage unemployment benefit claimants to participate in activities or training, but prior to ND referral take-up of such advice is voluntary.

- a) Preparation for Work at Level 1 or Below: All learners with a learning aim of 'Preparation for Life and Work' and/or 'Entry to Employment (E2E) pre-apprenticeship offer'¹¹ and/or 'Aims at Level 1 or Below'; and who do not have any higher FE learning aims. This is learning at European Qualification Framework (EQF) Level 2; it is at a level lower than that expected of the average [16-year-old] school leaver in England; and below the level of a US High-school Diploma. It is the most common form of learning amongst our unemployed cohort.
- b) Level 1/Level 2 Maths and/or English. All learners who have the highest, or only, learning aim of Level 1/Level 2 Maths; or Level 1/Level 2 English; or both. This is a group taking basic Maths and English at EQF Level 3 or Level 2; which is at or below the level of Maths/English achievement expected of the average [16-year-old] school leaver in England; or US High-school Diploma.
- c) Full Level 2 and above: This category includes more substantial learning aims that are equivalent to, or at a higher level than, EQF *Level 3*. In the UK this is equivalent to 5 GCSEs at grade A* to C or an NVQ2 (which is a technical/vocational qualification similar to the German Vocational Qualification Certificate, VQC). This is a category of learning that is mostly inhabited by those studying the equivalent of US Community College Certificates or German VQCs, but a small number study for the equivalent of US Associate Degrees/German Advanced VQCs.

Our categorization of learning into these groups is driven by consideration of (i) numbers (which must be sufficient to allow sensible econometric investigation) and (ii) the detail of information available in the ILR. Table 1 reports some basic descriptive statistics associated with these training schemes. All these data are then matched to information on employment and earnings in the WPLS.

		Mean	Duration	%	
Main		duration	(St.	Partially	
categories	Type of Training	(days)	Dev.)	Treated	Number enrolled
Preparation	Preparation for life and work	115	135	38%	16,144
for Work at	Entry to employment apprenticeship	306	254	44%	2,863
Below	Other programs at level 1 or below	96	102	39%	26,329
L1/L2 English and/or Maths	L1/L2- literacy and/or numeracy	57	105	37%	53,071

Table 1: Learning aims identified in the ILR

¹¹ E2E is a pre-apprenticeship 'offer' for those with few/no-qualifications and little experience. It covers basic reading, writing and communication skills; together with career planning and other basic employability skills.

Full Level 2 or above	Aims at level 2 or above of 120 Guided Learning Hours (GLH) or more	237	154	45%	9,592
	ICT Aims at level 2 or above less than 120 GLH/ unknown GLH	100	108	42%	1,472
	Other aims at level 2 or above less than 120 GLH/ unknown GLH	152	131	36%	12,998
	Aims at level 2 or above of 480 GLH or more	398	268	59%	842

Note: The number enrolled is limited to the highest aim attempted during the period before referral to the New Deal and also limited to those still unemployed at the time training started.

3.2Econometric Approach

Our framework of reference is similar to that of Sianesi (2008), where $D \in \{0,1,2,3\}$ represents the option of waiting or participation in one of our three categories of training. In this context, Y^d denotes the potential outcome an individual would experience were he/she to receive treatment d. We first assess the average effect for all programme d starters, compared to the waiting option:

(A) all starters and currently untreated: $E(Y^d | D = d) - E(Y^0 | D = d)$, for $d \in \{1, 2, 3\}$

While the first term of (A) is observed in the data, identifying assumptions need to be invoked to estimate all counterfactuals $E(Y^0 | D = d)$ - that is, the outcomes participants in d would have experienced, on average, had they chosen the waiting option. One key assumption is the CIA, which postulates that all relevant differences between treatment and control are captured in their observable attributes X. In this case, Y^0 represent a valid counterfactual scenario for individuals who received treatment d:

$$E(Y^0 | D = d, X = x) = E(Y^0 | D = 0, X = x), \text{ for } d \in \{1, 2, 3\}$$

Matching methods are usually adopted to produce a matched group in which the distribution of pre-treatment observables X is as similar as possible to the distribution in treatment group d. These methods additionally require that the common support assumption is satisfied, i.e. every treated individual is assumed to have at least one counterpart in the control group. When adopting a coarsened exact matching (CEM) approach, the use of more disaggregated categories to coarsen variables (i.e. using deciles as opposed to tertiles) makes it more challenging to find a match. A criterion used in our approach to exact matching, is that we lose less than 5% of the treatment group as a result of common support issues. The data utilised for this study provide us with large and heterogeneous groups of *currently untreated* and *partially treated*, and in the *Sensitivity Analysis* of Section 5 we confirm that our results are not sensitive to this assumption.

Despite the widespread use of several matching methods across disciplines (see Stuart, 2010), these methods have often been misapplied. Recently, a new class of matching methods has emerged - dubbed "monotonic imbalance bounding (MIB)" - that curtails the potential misuse of these techniques.¹² We implement one of these MIB methods, using coarsened exact matching (CEM) proposed by lacus et al. (2011; 2012). The detail of this approach is set out in Section 3.3.

However, our framework also utilises the information we have on achievement (or otherwise) of learning outcomes - splitting the group of all starters into *fully* and *partially* treated. Hence, letting FT denote the group of fully treated and PT the group of partially treated, we estimate employment impacts by also comparing:

(B) fully treated and currently untreated: $E(Y^{d,FT} D = d) - E(Y^0 D = d)$, for $d \in \{1,2,3\}$	
(C) partially treated & currently untreated: $E(Y^{d,PT} D = d) - E(Y^0 D = d), d \in \{1,2,3\}$	for
(D) fully treated and partially treated: $E(Y^{d,FT} D = d) - E(Y^{d,PT} D = d)$, for $d \in \{1,2,3\}$	

As suggested in Section 1, the identification strategy under comparison (A) will only remain valid for comparisons (B), (C) and (D) under the assumption that any unobserved factors driving selection into FT and PT are uncorrelated with potential outcomes, having conditioned on observed learning and employment histories prior to treatment start. Our ability to include detailed employment and learning histories, including indicators of previous partial/full treatment, helps in this respect. However, studies by Kluve et al. (2012), who find the duration of training to be endogenous, and Fitzenberger et al. (2015) who account for endogenous selection into dropout, question the validity of this assumption. If there are unobserved factors correlated with potential outcomes, that also drive selection into partial/full treatment, and this information only arises following treatment start, we are at risk of bias in our comparison of effects from full and partial treatment. Whilst Section 2.1 describes our focus on measures of partial and full treatment that differ from the 'dropout' and 'duration' analysed in these studies, the same potential exists for comparison of outcomes across the fully and partially treated to be confounded by [unobserved] nonrandom selection into these states. Therefore, in Section 5.1 we investigate this issue of selection following programme start, using distance to nearest FE training provider to instrument for possible endogenous selection into partial and full treatment.

Finally, in our study where the timing of training initiation is voluntary and can vary across individuals, a static approach to evaluation would ignore the dynamic element of selection into treatment. If we simply compare all unemployed individuals initiating a treatment over

¹² Differently from the most widely used matching methods, such as propensity score and Mahalanobis matching, the MIB class of matching methods is able to explicitly bound model dependence. MIB implies that relaxation of one tuning parameter for one variable, controls monotonically the imbalance measures, without altering the maximal imbalance on the remaining variables (lacus et al., 2012).

the 18-month period, with those who do not, we are selecting the comparison group conditional on them never being treated over that period. As a result, unemployed individuals may be included in the comparison group simply because they secure employment prior to any potential training initiation and this amounts to conditioning on future outcomes. Therefore, in each of our analyses we separately estimate employment impacts for (a) those who initiate training in the first two months of their unemployment spell; (b) those who first initiate training in the third or fourth months of unemployment; (c) those who first initiate training in the first time in the seventeenth and eighteenth months from claim start.

Under (A) the comparison is therefore between all those who start a training treatment at a point in time [all starters], and those whom we do not observe initiating a treatment [currently untreated] up to that specified point in time, but who may do so in the future, up to 18 months from claim start date. Estimands from analysis under (B) and (C) are obtained using a similar approach, as the causal contrast is between 'training now' (whether this is for all starters, the fully treated or partially treated at a point in time) versus 'waiting' (i.e. those who at the specific point in time, are *currently untreated*). However, estimands obtained from (D) differ to those under (A), (B) and (C) as unemployed individuals who start training during a given sub-interval and achieve [fully treated], are compared to the unemployed who start training during the same sub-interval, but do not subsequently achieve the learning outcomes [partially treated]. Following Sianesi (2008), we present figures that are aggregates of these effects estimated by time of initiation, weighted according to the observed distribution of initiation across the nine separate periods. Under (A), (B) and (C) Sianesi (2008) suggests that, whilst "a clear casual interpretation only pertains to the effects by month of entry, this summary measure offers a synthetic overview of the general patterns of the effects of [training] by month of placement"¹³. In contrast, estimates obtained from (D) are closer to those used in standard cost-benefit frameworks.

3.3 The detail on Coarsened Exact Matching

The idea of CEM is to 'coarsen' each conditioning variable into meaningful categories¹⁴, and then match exactly on these 'coarsened' variables to determine the matches and to 'trim' unmatched units. Finally, the coarsened data are discarded and the original (un-coarsened) values of the matched data are retained. Therefore, the CEM algorithm creates a set of strata, each with the same coarsened values of the set of conditioning variables. If different numbers of treated and control units appear in different strata, the econometric model must weight or adjust for the different stratum sizes. This is why a weighted regression of the dependent variable on the covariates is adopted at the end of the matching procedure. Combining CEM with a weighted regression helps to control for any remaining mismatch between treated and comparison individuals. Iacus et al. (2011) show that CEM outperforms commonly used existing matching methods in its ability to reduce imbalance, model dependence, estimation error, bias, variance, mean square error, and other criteria in

¹³ Versions of Figures 1, 2 and 3 created using a 'static' approach to matching are available from the authors on request and these estimates of the average treatment effect on the treated (ATT) do not deviate substantially from the figures presented here for (A), (B) and (C).

¹⁴ For instance, if we are matching on previous earnings, we may match on data that has been 'coarsened' by putting earnings into quartiles.

several real and simulated data sets. However, using CEM, the coarsened values are chosen by the researcher in a customized way based on substantive knowledge of the measurement scale of each variable. Therefore, in Section 5.2 we check whether the estimates substantially change when using different coarsened values.

Our analyses plot the pairwise percentage point (ppt) difference in employment of *all starters*, *fully treated*, *partially treated* and *currently untreated*, before the claim start date and after the programme start date, having:

- Matched (using CEM) on pre-programme unemployment duration, number of months in employment between month (t-1) and month(t-60) split in quintiles, number of months in employment between month (t-61) and month (t-96) split at the median, number of months in active benefits between month (t-1) and month (t-60) split in quintiles, referral of a caseworker¹⁵, sex, age group split in three groups, white/non-white ethnicity, disability dummy, local unemployment rate split at the median, and number of prior ILR aims started split at the median.¹⁶
- We then follow the evaluation procedure suggested by Blackwell et al. (2009) after obtaining the CEM weights, we use a weighted regression to adjust for the imbalance which remains in the matched data after coarsening¹⁷. In each weighted regression we control for the aforementioned variables as well as for whether an unemployed individual has children, main ethnic groups, whether individual is a previous offender, age, ever lone parent, ever asylum seeker, language capabilities, postcode district fixed effects, local index of multiple deprivation, annual earnings in year t-1, number of prior LMS opportunities, prior mandatory referrals, and number of prior ILR aims started and achieved.

A key test of whether we achieve such balance in the characteristics of our treatment and comparison groups, as Heckman et al. (1998b) suggest, is to show that pre-treatment employment rates of the two groups are similar over the previous 8 years. If this is the case, we can be more confident that any subsequent divergence in employment rates at the point of treatment, is not due to differences in time-invariant unobserved factors. In Section 4 we report a number of figures showing that, after the CEM procedure, there is no statistically significant evidence of differences in employment trends prior to the relevant unemployment spell for all pairwise comparisons. A recent paper by Caliendo et al. (2017) confirms the importance of conditioning flexibly on lagged employment and wages, benefit receipt history, and local labour market conditions (also see Heckman and Smith, 1999; Lechner and

¹⁵ Differently from Sianesi (2008) and Biewen et al. (2014), we do not have detailed job-seeker profiles, as reported by caseworkers. However, we do know if the caseworker considered the unemployed individual to be in need of basic support and guidance (as flagged in the LMS) and we exactly match treated and Currently Untreated unemployed on this.

¹⁶ We coarsen on the most relevant variables highlighted by the ALMP literature. This is in line with lacus et al. (2012: page 16) who recommend that "users of CEM choose the coarsening criteria based on their knowledge of the covariate measurement process and other substantive criteria such as the likely importance of different variables". We then opt for a coarsening specification quite detailed but which, at the same time, allows us to keep at least 95% of treated observations in the common support for all comparisons.

¹⁷ Selecting matched samples reduces bias due to covariate differences, and regression analysis on those matched samples can adjust for small remaining differences and increase the efficiency of our estimates (Stuart and Rubin, 2007).

Wunsch, 2013)¹⁸, by showing that additional variables (not usually observed in studies of ALMP) do not change estimates when added to the conditioning set.

4. Estimating training impacts for the fully and partially treated

Figures 1 to 3 present findings from estimation of training impacts, associated with [1] *Preparation for Work at Level 1 or Below* (learning at or below European Qualification Framework, EQF *Level 2*); [2] *Level 1/Level 2 Math and/or English* (Maths and English at EQF *Level 2* or *Level 3*) and [3] *Full Level 2 and above* (more substantial learning aims, most of which are equivalent to EQF *Level 3*). Each Figure contains four pairwise comparisons: [A] *all starters* and *currently untreated*; [B] *fully treated* and *currently untreated*; [C] *partially treated* and *currently untreated*; and [D] *fully treated* and *partially treated*. The *fully treated* and *partially treated* are defined as such according to whether the [ILR] educational administrative dataset, records that they do, or do not, achieve learning outcomes, having enrolled for training. The estimates for each of these pairwise comparisons are obtained using CEM, with the impacts at each month from treatment start averaged across nine separate analyses, to accommodate the potential for dynamic selection into training start¹⁹.

Adopting Sianesi's (2008) approach to the presentation of this analysis, we aggregate all estimated treatment effects by the time of initiation and weight according to the observed distribution of initiation across the nine periods. This allows us to create figures that clearly show the evolution of impacts over time, with the vertical axis measuring the percentage point (ppt) difference between the observed employment percentage amongst [in the case of Figure 1A] *all starters* and the estimated counterfactual percentage. On the horizontal axis, positive values denote months since course start, while negative values represent pre-unemployment months.

The period between t-8 years and claim start date (t=0) provides a graphical indication of the quality of the match. In Figure 1A, the thickened line representing the ppt difference remains close to zero for the entire period up to claim start date and well within our 99%

¹⁸ The Stable Unit Treatment Value Assumption [SUTVA] assumes that (i) treatment applied to one unit, does not impact other units and (ii) there is one version of each treatment (Rubin, 1986). Under (ii), we group similar types of learning, but variety remains within our category of, for instance, *Preparation for Work*; not least because it will be delivered by a variety of trainers. However, this would seem to be less acute than in existing ALMP studies, where there is a question of what constitutes training (McCall et al., 2016). Under (i), our study adopts the assumption, implicit in many econometric investigations, that potential outcomes are independent across individuals, ruling out general equilibrium effects. Again, we cannot dismiss these concerns, but feel that general equilibrium effects such as displacement are less of an issue given the nature of treatments analysed. The majority of training we consider is at a level expected of the average school leaver and is therefore 'remedial' in nature – as a result, the increased supply of skills to the market cuts across a wide variety of professions and may be argued to lessen the likelihood of displacement. These considerations do not wholly rule out the potential for increased supply of these skills amongst the treated, to impact job opportunities of the comparison group; but we argue that they do place limits on such concerns (see Greenberg et al. 2011; McCall et al., 2016: Section 2.7, for a more detailed discussion), when compared to higher-level learning, which tends to become more specific to a particular group of occupations.

¹⁹ The full set of underlying point estimates and standard errors, for the sub-intervals and the aggregates that appear in Figures 1 to 3, are available from the authors on request.

confidence interval, obtained using robust standard errors²⁰ (the light dashed lines on either side). For each of our estimated training treatments, percentage point (ppt) employment impacts are associated with overall employment rates of around 30 per cent to 40 per cent in the years after learning (that is, an approximate 35% employment rate on average).

The main finding from Figure 1, Panel A [from here, 'Figure 1A'] is the large and persistent statistically significant impact of *Preparation for Work* on the employment prospects of unemployed individuals aged 25 to 55. In Panel A we observe an employment premium for all starters, over the currently untreated aged 25 to 55 years old, which rises to just under 6 percentage points [ppts] six months after training initiation and remains above 5 ppts at a point three years from training initiation. This approximate 5 ppt employment effect is equivalent to an approximate 14% impact, in the context where we observe employment rates of around 35 per cent.

Figure 1: Employment impacts comparing across 'all starters', 'fully treated', 'partially treated' and 'currently untreated' [Preparation for Work at Level 1 or Below]



A: All Starters and Currently Untreated

B: Fully Treated and Currently Untreated

²⁰ We have also carried out the analysis using bootstrapped standard errors, and this leads to very similar estimates.



C: Partially Treated and Currently Untreated



D: Fully Treated and Partially Treated



Note: The solid line represents the pointwise difference in employment rates, while the dashed lines are pointwise 99% confidence intervals. There are 30,123 treatment *starters* (18,875 *fully* treated and 11,248

partially treated)²¹ and the '*currently untreated*' group includes 570,057 individuals who do not initiate treatment during the period of analysis.

Whilst we can draw on existing evidence (Heckman and Smith, 1999; Lechner and Wunsch, 2013; and Caliendo et al. 2017), to invoke the CIA for comparisons in Panel A; there is currently no evidence to support our assumption that any unobserved factors driving selection into *partial/full* treatment are uncorrelated with potential outcomes, having conditioned on observed learning and employment histories prior to treatment start. This assumption is required for credible invocation of the CIA when considering the results in Panels B, C and D; and in Section 5.1 we investigate further using distance to nearest FE training provider to instrument for possible endogenous selection into partial and full treatment, following treatment start. Findings from Section 5.1 suggest that the potential for endogenous selection following treatment start apparent in existing studies (Kluve et al., 2012; Fitzenberger et al., 2015), does not invalidate the comparisons made in Panels B, C and D.

Given this support for our identification strategy, we can be more confident in the pattern of impacts across Panels B to D that suggest *partially* treated individuals secure some form of return to human capital accumulation, and this is on average at a lower level than that achieved by the *fully* treated. Panel B of Figure 1 reports a similar impact to that of Panel A, with the average impact of training remaining at or around 6ppts until a point 3 years on from training initiation, when we compare the *fully* treated with the *currently untreated*. Panel C identifies an employment premium for the *partially* treated, when compared to the *currently untreated*, of approximately 4 ppts between 3 months and four years from training initiation; whilst Panel D suggests an employment premium for the *fully* treated over the *partially* treated that varies around the 2 ppt point level from a point 6 months from training initiation. The findings from Figure 1 suggest a hierarchy of impacts consistent with those found in existing studies (for instance, Bell et al., 1995; Heckman et al., 1998a; Paul, 2015; Choe et al., 2015), where the *partially treated* are seen to secure a return to human capital accumulation [over the *currently untreated*] that is statistically significant, but less than that secured by the *fully treated*.

Figure 2 presents estimated employment premiums that arise from comparisons of *all starters*, *fully* and *partially* treated unemployed individuals aged 25 to 55 who engage in *Level 1/Level 2 Maths &/or English* learning. Here, Panel A identifies an employment premium for *all starters*, over the *currently untreated*, which rises to 4 percentage points [ppts] six months after training initiation and remains above 2 ppts at a point three years from training initiation. Panel B suggests that an employment premium for the *fully treated*, compared to matched *currently untreated* unemployed individuals aged 25 to 55, rises quickly to around 4 ppts following training initiation and remains at that level four years later. Panel C of Figure 2 confirms this hierarchy of impacts, with the *partially treated* securing an

²¹ Of the 30,123 treatment starters, 5,547 initiated training in the first two months of their unemployment spell; 4,708 in the third or fourth months; 3,818 in the fifth or sixth months; 3,074 in the seventh or eight months; 2,778 in the ninth or tenth months; 2,538 in the eleventh or twelfth months; 2,546 in the thirteenth or fourteenth months; 2,626 in the fifteenth or sixteenth months; and 2,488 in the seventeenth or eighteenth months.

employment impact (over the *currently untreated*) that remains between 2 and 3 ppts up to 24 months from training initiation (but which does not persist beyond 36 months). Panel D suggests that the *fully treated* secure an employment premium over the *partially* treated (who initiate training during the same sub interval) of around 2 ppts between six and 36 months from initiation of training.

Figure 2: Employment impacts comparing across 'all starters', 'fully treated', 'partially treated' and 'currently untreated' [L1/L2 Maths and/or English]



A: All Starters and Currently Untreated

B: Fully Treated and Currently Untreated



C: Partially Treated and Currently Untreated







Note: The solid line represents the pointwise difference in employment rates, while the dashed lines are pointwise 99% confidence intervals. There are 36,037 treatment *starters* (23,104 *fully* treated and 12,933 *partially* treated)²² and the '*currently untreated*' group includes 566,397 individuals who do not initiate treatment during the period of analysis.

²² Of the 36,037 treatment starters, 5,780 initiated training in the first two months of their unemployment spell; 4,943 in the third or fourth months; 4,569 in the fifth or sixth months; 4,325 in the seventh or eight months; 3,844 in the ninth or tenth months; 3,143 in the eleventh or twelfth months; 3,160 in the thirteenth or

The findings from Figures 1 and 2 imply that an 'intention to treat' approach to the estimation of impacts [as adopted in Panel A], would not alter our conclusions regarding the overall efficacy of training interventions considered here. Our identification of an effect from *partial* treatment that is close to that for the *fully* treated, is reassuring in this sense when considering the approaches that dominate in the ALMP literatures. In contrast, studies in the education economics literature that attempt to capture returns to vocational learning using the partially treated to estimate counterfactual outcomes, risk understating impacts (Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Jepsen et al., 2014; Bibby et al. 2014; Bibby et al., 2015; Hedges et al., 2018). If there are strong selection effects from the nontreated to the treated, and these are difficult to credibly deal with via conditioning on observed variables, the partially treated may provide a useful comparison, as we observe them enrolling for (selecting into) a course, but not receiving full treatment. However, Figures 1 and 2 suggest that estimates obtained from such comparisons may understate impacts.

Finally, Figure 3 sets out the estimated impacts accruing to 25 to 55 year olds undertaking more substantial training interventions at *Full Level 2 and above*. In Panel A of Figure 3, we observe an employment premium for *all starters* over the *currently untreated* that rises above 6 ppts twelve months from initiation of training and remains above 6 ppts until 40 months from training initiation. As was the case in Figure 1D, we observe a lock-in effect when comparing the *fully* and *partially* treated in Figure 3D and this is statistically significant up to a point around six months from training initiation. Following this period of lock-in, the *fully* treated secure an approximate 2 percentage point premium over the *partially* treated between 14 and 44 months from training initiation. Whilst Panel B of Figure 3 implies some lock-in when comparing *fully* treated with the *currently untreated*, this is not statistically significant; and in this panel the main finding is of an employment premium that rises above 6 ppts twelve months from initiation of training and remains between 7 and 8 ppts, in the period twenty to forty-six months from training initiation.

As was the case for Figures 1 and 2, Panel C of Figure 3 confirms the pattern of effects we might expect from a growing ALMP literature (for instance, Bell et al., 1995; Heckman et al., 1998a; Paul, 2015; Choe et al., 2015) that identifies impacts accruing to dropouts (in our context, a subset of the *partially treated*) – assuming the set of conditioning variables is sufficient to justify invocation of the CIA for comparisons in Panels B to D. In Panel C the *partially treated* are estimated to secure an impact of around 6 ppts over the *currently untreated* that is statistically significant; but less than the 7 to 8ppts secured by the *fully treated*, when compared to the *currently untreated*.

In addition, Figure 3 repeats a pattern evident in Figures 1 and 2, which suggests impacts from *partial* treatment exhibit less persistence, when compared to those accruing to the *fully* treated. In each case, the decline in estimated impacts for *all starters* [evidenced in Panel A] seems driven by the lesser persistence of impacts accruing to the *partially* treated [Panel

fourteenth months; 3,369 in the fifteenth or sixteenth months; and 2,903 in the seventeenth or eighteenth months.

C] when compared to those of the *fully* treated [Panel B]. This seems most pronounced when considering Panel C of Figure 2 but is also apparent in the other Figures.





A: All Starters and Currently Untreated

B: Fully Treated and Currently Untreated





D: Fully Treated and Partially Treated



Note: The solid line represents the pointwise difference in employment rates, while the dashed lines are pointwise 99% confidence intervals. There are 12,273 treatment *starters* (7,536 *fully* treated and 4,737 *partially* treated)²³ and the *'currently untreated'* group includes 565,954 individuals who do not initiate treatment during the period of analysis.

The lack of substantial lock-in effects across Figures 1, 2 and 3 is perhaps unsurprising, given the focus of our analysis. First, as Wunsch (2016: page 4) suggests, 'lock-in effects are expected to be moderate for a range of shorter programs' and the majority of training considered in this study is of a short duration. For example, a *Functional Skills qualification in English at Entry Level 1* [a specific qualification included within the category of *L1/L2 English and/or Maths*] has 45 guided learning hours [GLH] mandated by the awarding body, and depending on the specific college delivering the learning, this will be between 1 and 2 days a week for approximately 5 weeks²⁴.

²³ Of the 12,273 treatment starters, 1,613 initiated training in the first two months of their unemployment spell; 1,481 in the third or fourth months; 1,398 in the fifth or sixth months; 1,239 in the seventh or eight months; 1,256 in the ninth or tenth months; 1,237 in the eleventh or twelfth months; 1,427 in the thirteenth or fourteenth months; 1,348 in the fifteenth or sixteenth months; and 1,273 in the seventeenth or eighteenth months.

²⁴ More generally, one can see that the category of *L1/L2 English and/or Maths* is associated with a mean duration of 57 days. This is the average number of days from the point of course commencement to completion/dropout, as recorded by the ILR administrative system. In the case of a *Functional Skills*

Wunsch (2016: page 4) flags, 'the employment prospects of participants if they were not enrolled in the program', as the second most important determinant of lock-in size. Table A2 of the Appendix shows that for those undertaking *Preparation for Work* (which includes training of a longer average duration than that included in the category of *L1/L2 English and/or Maths*) the average age of participants is just under 41 years [across the *fully treated*, *partially treated* and *currently untreated* groups]. The job prospects of these older individuals, whose literacy and/or numeracy is substantially below that expected of the average 16-year-old school leaver, are extremely limited in the absence of programme participation. When we consider the higher categories of learning at *Full Level 2 and above*, Table A2 suggests we are dealing with less challenging groups. But the difference is only small, as the average age is still 36 years, 23% have indicators of disability and the majority of this training is at the level we expect of the average school leaver.

The short duration of many interventions; very low employment probabilities in the absence of intervention; and voluntary nature of the training considered²⁵ all provide explanations for the limited amount of lock-in. The only statistically significant lock-in effects are observed for *Preparation for Work* (where we have comparison groups with very limited employment prospects, but where durations are longer) and *Full Level 2 and above* (where durations are longer, and the comparison group are potentially more employable). However, even here the lock-in effect is only significant when comparing the fully treated and partially treated, reflecting the very limited employment prospects for those who do not engage in an intervention.

As already suggested, the credibility of our findings rest on demanding assumptions regarding our ability to match up to the point of training initiation, in a way that accommodates potentially endogenous selection into *partial* and *full* treatment following training start. Section 5.1 now considers the validity of this assumption, presenting an analysis that instruments for possible endogenous selection into these states. Before considering this analysis, it is important to note the results of our analysis carried out for the younger 18 to 24 age group, using a similar approach to the identification of impacts. For this younger age group, the period prior to mandatory New Deal referral is much shorter [6 months] but the approach to analysis is the same, using CEM in the Sianesi (2008) framework, to justify invocation of the CIA.

The estimates for this younger age group are presented in Appendix Table A1, and as a brief review of the findings suggests, we obtain a similar pattern of results to those for the 25 to 55 age group when considering the *Full Level 2 and above* learning category but for our other two categories of learning, this is not the case. It is important to note concerns over data limitations, when considering the 18 to 24 age group. Findings from Heckman and

qualification in English at Entry Level 1 delivered over a five week period, the duration recorded in the administrative system would approximate 40 days [5 weeks from start date to finish date, plus approximately 5 days to account for the short period between official registration and learning start].

²⁵ Take-up of training is voluntarily during the period under study (i.e. prior to any mandatary ND referral) and whilst case workers may be less stringent in their insistence on job-search during the programme of study, the individual still counts as unemployed for case-worker performance. Furthermore, there is no incentive to stay on the programme as benefits are not impacted either way. Wunsch (2016) notes that these characteristics are also likely to reduce the size of any lock-in effects.

Smith, (1999); Lechner and Wunsch, (2013); and Caliendo et al. (2017) suggest that the lack of a detailed labour market history for many in this age group limits our ability to credibly invoke the CIA, even when considering comparisons of [A] *all starters* with the *currently untreated*. Existing studies tend to concentrate on older age groups, aged 25 years and above (Fitzenberger et al., 2006; Sianesi, 2008; Dolton and Smith, 2011; Biewen et al., 2014) and even when studies analyse individuals from a broader age range (for instance, Sianesi 2004; Lechner et al., 2011) they tend not to capture impacts specifically for this younger age group. Our study would suggest that identification of robust causal impacts for this younger age group remains a challenge that may not be overcome using the identification strategies adopted here.

Overall, the findings in this section suggest an important role for training at levels of learning between EQF Level 1 and Level 3, for unemployed individuals aged 25 to 55 who in most cases are attempting to remediate poor levels of attainment that have persisted since the age of 16. The focus of discussion in Section 6 is on the additional insights that consideration of the *fully* and *partially* treated provide; and the extent to which our identification strategy accommodates the challenge of possible endogenous selection following treatment start, is now investigated in Section 5.

5. Sensitivity Analyses

5.1 Endogenous Selection into Full/Partial Treatment

As already suggested, the credibility of findings in Figures 1 to 3 rest on demanding assumptions regarding our ability to match up to the point of training initiation, in a way that accommodates potentially endogenous selection into partial and full treatment, following training start. The Instrumental Variable (IV) approach used in this section is based on the two-stage least squares (2SLS) estimator and it requires a variable that influences the probability of attending training, as well as the probability of selection into partial/full treatment, but which does not influence employment outcomes. The intuition behind this is that the instrument can be viewed as introducing essentially random variation in participation in a similar way to a randomised controlled trial (Dorsett et al., 2019). If this holds, identification of causal impacts no longer relies on explicitly controlling for all important differences between the *fully*, *partially* and *currently untreated*.

Our instrument is the distance from an unemployed individual's place of residence, to the nearest FE training provider. Distance to college was first used as an instrument for schooling by Card (1995) and more recent applications include the estimation of impacts arising from traineeships (see Dorsett et al., 2019). We construct a continuous measure of typical travel distance by car between the postcode district centroid in which the individual resided at the beginning of the unemployment spell and the closest FE training provider. This measure is the one used to obtain findings presented in Table 2 and an additional sensitivity test confirms these findings when we utilise an alternative geodesic ('as the crow flies') measure of distance.

Travel distance is seen as a cost-shifter – the larger the distances to an FE training provider, the higher the transaction cost of engaging in education and training. Studies have shown that distance to college is an important determinant of enrolment and much of this is due to

the associated variation in transaction costs (Frenette, 2004; Spiess and Wrohlich, 2010). Here we are using distance to instrument for both training enrolment and completion (as captured by indicators of *full/partial* treatment), as it is considered a cost-shifter for both. In much of the literature (for instance, Walters, 2018) distance is used to instrument for enrolment; but in the education economics literature, it is not uncommon to instrument for the endogeneity of educational completion (see for instance, Dickson 2009).

Existing studies have challenged the use of distance to college as a valid instrument, as in some contexts it is correlated with indicators of student ability and family background characteristics (see Kjellström and Regner, 1999; Cameron and Taber, 2004). This concern arises in situations where individuals/families have incentives to locate closer to high-quality schools/colleges that form the focus of evaluation. Where this is the case, we would expect to observe more affluent families and/or more able students having shorter travel distances. In the study here, we are highly unlikely to observe such behaviours, as Further Education has the lowest levels of funding per pupil across the English education system (Belfield, Crawford and Sibieta, 2018); and the majority of learning treatments undertaken by unemployed individuals aged 25+ are at or below the level expected of an average [16-year-old] school leaver.

Our identification of employment impacts in Table 2 is therefore based on the assumption that distance is predictive of engagement in FE training; but is not correlated with unobserved factors influencing employment outcomes. To support this, we find the Kleibergen-Paap rk Wald F-statistics associated with the first stage regressions to range from 81.2 to 261.7 and this satisfies the standard weak instrument tests. Also, the coefficient estimates for the excluded instrument (the inverse log distance from the closest FE training provider) are always positive and statistically significant at the 1% or 5% level; indicating that living closer to an FE training provider is associated with an increased probability of training enrolment, as well as a reduced probability of selecting into partial treatment.

The estimates are obtained by first computing the CEM weights as in the main analysis and then by employing the weighted 2SLS estimator. Table 2 presents the estimates. Overall the findings support those obtained from approaches to matching set out in Figures 1 to 3. Our use of a continuous instrument means the 2SLS estimator can be considered as a weighted average of Local Average Treatment Effects [LATE]²⁶, with the overall IV effects presented in Table 2 being representative of all compliers for changes across all values of the instrument (Cornelissen et al., 2016: Section 2.3). Recent developments in labour economics emphasise the potential for estimation of Marginal Treatment Effects (MTE) as a more informative way to exploit a continuous instrument such as ours (for instance, Carneiro, Heckman, and Vytlacil, 2011), but this takes us beyond the scope of the current paper²⁷.

Estimates presented in Table 2 are aggregated across a greater number of time periods than those presented in Figures 1 to 3, but this only impacts comparison over the first 12 months following training initiation. For instance, when considering *A: All Starters &*

²⁶ As Cornelissen et al. (2016) suggest, the weight given to each of these LATEs in creation of the overall average is positively related to (i) the first stage of the 2SLS process, with 'a stronger first stage getting a higher weight' and (ii) the 'group size' associated with estimation of each LATE.

²⁷ See Cornelissen et al. (2016) for an excellent discussion of the issues around estimation of MTE.

Currently Untreated in the learning category of *Preparation for Work* in Table 2, the first estimated employment ppt impact is 0.036 at 12 months. This figure of 0.036 is an average across these first 12 month and corresponds to a figure of 0.049 if we take the same approach to presentation of the impact accruing during the first 12 months in Figure 1 Panel A. Across Table 2 there is some tendency for estimates in these first 12 months to be slightly lower using the IV approach, when compared to those in Figure 1 to 3; and for estimates over the final fourth year to be slightly higher. However, over the entire period under study, it would seem reasonable to suggest that Table 2 provides support for our invocation of the CIA across Figures 1 to 3, even when considering selection into *partial/full* treatment.

Percentage po	pint employme	ent gap in years	s after learning	spell start
	1 st Year	2 nd Year	3 rd Year	4 th Year
A: All Starters	0.036***	0.050***	0.056***	0.069***
& Currently Untreated	(0.006)	(0.007)	(0.007)	(0.008)
B: Fully Treated & Currently Untreated	0.042*** (0.007)	0.060*** (0.008)	0.067*** (0.009)	0.081*** (0.009)
C: Partially Treated & Currently Untreated	0.026*** (0.008)	0.033*** (0.010)	0.037*** (0.011)	0.030*** (0.011)
D: Fully Treated & Partially Treated	0.018* (0.0011)	0.027** (0.011)	0.022* (0.013)	0.037*** (0.012)
A: All Starters & Currently Untreated	0.021*** (0.006)	0.028*** (0.006)	0.036*** (0.007)	0.033*** (0.007)
B: Fully Treated & Currently Untreated	0.025*** (0.007)	0.041*** (0.009)	0.048*** (0.008)	0.039*** (0.010)
C: Partially Treated & Currently Untreated	0.020** (0.008)	0.015 (0.010)	0.003 (0.011)	0.010 (0.012)
D: Fully Treated &	0.012 (0.010)	0.031*** (0.010)	0.034*** (0.012)	0.040*** (0.013)
Partially Treated				
	Percentage por A: All Starters & Currently Untreated B: Fully Treated & Currently Untreated C: Partially Treated & Currently Untreated D: Fully Treated & Partially Treated & Partially Treated & Currently Untreated B: Fully Treated & Currently Untreated Currently Untreated D: Fully Treated & Currently Untreated Currently Untreated D: Fully Treated & Currently Untreated C: Partially Treated & Currently Untreated C: Partially Treated & Currently Untreated C: Partially Treated & Currently Untreated & Currently Curently Cu	Percentage point employme1st YearA: All Starters0.036***& Currently(0.006)B: Fully0.042***Treated &(0.007)Currently0.026***Untreated(0.008)C: Partially0.026***Treated &(0.008)Currently0.018*Treated &(0.0011)Partially0.021***A: All Starters0.021***& Currently(0.006)Untreated0.025***Freated &(0.007)Currently0.025***Treated &(0.007)Currently0.020**Treated &(0.008)C: Partially0.020**Treated &(0.008)C: Partially0.020**Treated &(0.008)CurrentlyUntreatedD: Fully0.012Treated &(0.0010)	Dercentage point employment gap in years 1 st Year 2 nd Year A: All Starters 0.036*** 0.050*** & Currently (0.006) (0.007) Untreated 0.042*** 0.060*** B: Fully 0.042*** 0.060*** Treated & (0.007) (0.008) Currently 0.026*** 0.033*** Treated & (0.008) (0.010) Currently 0.026*** 0.033*** Treated & (0.008) (0.010) Currently 0.018* 0.027** Untreated 0.011) (0.011) P: Fully 0.018* 0.027** Treated & (0.0011) (0.011) Partially 0.028*** (0.006) M: All Starters 0.021*** 0.028*** & Currently (0.006) (0.006) Untreated 0.025*** 0.041*** Treated & (0.007) (0.009) C: Partially 0.020** 0.015 Treated &	Percentage point employment gap in years after learning 1*t Year 2 nd Year 3 rd Year A: All Starters 0.36^{***} 0.050^{***} 0.056^{***} & Currently (0.006) (0.007) (0.007) Untreated (0.006) (0.007) (0.007) B: Fully 0.042^{***} 0.060^{***} 0.067^{***} Treated & (0.007) (0.008) (0.009) Currently 0.026^{***} 0.033^{***} 0.037^{***} Treated & (0.008) (0.010) (0.011) Currently 0.026^{***} 0.033^{***} 0.037^{***} Treated & (0.008) (0.010) (0.011) Currently 0.018^{*} 0.027^{**} 0.022^{*} Treated & (0.0011) (0.011) (0.013) Partially Treated & (0.006) (0.007) Untreated 0.021^{***} 0.028^{***} 0.036^{***} & Currently (0.007) (0.009) (0.008) </th

Table 2: Employment impacts estimated using the IV approach, compared across 'all starters', 'fully treated', 'partially treated' and 'currently untreated'

& Currently Untreated	(0.007)	(0.007)	(0.008)	(0.008)	
B: Fully Treated & Currently Untreated	0.002 (0.008)	0.024** (0.009)	0.053*** (0.010)	0.063*** (0.010)	
C: Partially Treated & Currently Untreated	0.044*** (0.010)	0.047*** (0.010)	0.040*** (0.010)	0.032*** (0.011)	Note: Robust standard errors using the Huber–
D: Fully Treated & Partially Treated	-0.020* (0.011)	-0.008 (0.012)	0.021 (0.014)	0.030** (0.015)	White variance- covariance matrix are reported in
					•

parentheses for all estimates.

***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.

5.2 Further Sensitivity Analyses

Table A2 presents any differences in the average pre-treatment characteristics of the *fully treated*, compared to both the *partially treated* and *currently untreated*; even after we obtain a match on prior employment histories. Table A2 suggests some remaining differences when we consider whether an individual (a) was ever a lone parent, (b) had any prior ILR starts and (c) had any prior ILR achievement. When considering these three indicators, there is a slight tendency for the *fully treated* group to have fewer individuals with a prior lone parent indicator; and more individuals with prior ILR starts and achievement. There is also some tendency for the *fully treated* to contain a lesser proportion of prior offenders. In compensation, we find that the *fully treated* have a higher IMD score, which reflects a higher average level of local deprivation.

To test the extent to which these continuing small imbalances could be influencing our findings, we present estimates in Table A3 which use Propensity Score Matching (PSM) [after having dropped the observations falling off the common support using CEM] rather than a weighted regression.²⁸ In Table A4 we additionally test the sensitivity of our estimates by adopting Inverse Probability Weighting (IPW) which is a weighting estimator, as it has good statistical properties (such as attaining the semi-parametric bound under certain conditions) and good Monte Carlo performance (see e.g. Huber et al., 2013). Both sensitivity checks show that our findings remain largely unchanged.

Table A5 then repeats the analysis using sustained employment (with an individual counted as being in sustained employment if we observe them employed in that month, as part of a period of continuous employment lasting at least 6 months, with no overlapping active benefit spells). Again, this perspective does not alter our overall findings and the sustained employment premiums are very close to those we observe for employment outcomes in the main body of the paper. This is likely due to the fact that the benefits system does not allow

²⁸ We have used the k-nearest neighbor matching with n=10 and caliper=0.002. We have obtained similar estimates by trying other k-nearest neighbor matching specifications: i) n=10 and no caliper; ii) n=1 and caliper=0.002; iv) n=1 and no caliper.

'frictionless' movement between unemployment and employment, so one would expect claimants to mainly move into employment when they feel they have a job opportunity that will be sustained.

We focus on training programmes for the unemployed undertaken during the earlier months of an unemployment spell, when we are likely to observe few mandatory training events – though we would expect that much of the training is undertaken following a referral from a Jobcentre Plus advisor. It is quite possible that the decisions of some individuals over whether to participate in training during this earlier period of unemployment, are influenced by their expectation of interventions during the latter period when interventions are mandated [following 18 months for those aged 25 to 55]. Table A6 repeats the estimates presented in Section 4, focusing just on the sub-intervals in the first 12 of the 18 months prior to "mandatory" New Deal services. Again, our findings remain largely unchanged.

As highlighted in Section 3.3, the coarsened values are chosen by the researcher in a customized way based on substantive knowledge of the measurement scale of each variable; therefore, we now check whether the estimates substantially change when using different coarsened values. In Table A7 we report the estimates obtained using larger bins (less coarsening).²⁹ This analysis demonstrates that our estimates are robust to modest changes in the coarsening choices.

6. Summary of Findings and Implications for Future Research

This study considers training for the unemployed undertaken during the first 18 months of an unemployment spell, when training treatments are voluntary in nature. Large numbers in both treatment and comparison groups allow us to use Coarsened Exact Matching (CEM), to create month-by-month estimates of employment effects, up to 5 years from training start. We match [via CEM] on a variety of characteristics, including socio-demographic variables, prior qualifications, a flag of basic skill need from a caseworker and up to 8 years of labour market and learning histories. The potential for variation across individuals in the timing of treatment, raises concerns over the issue of dynamic selection into treatment. Therefore, we adopt the approach of Sianesi (2008), presenting figures that are aggregates of treatment effects estimated by time of initiation, weighted according to the observed distribution of initiation across periods.

Our substantive contribution arises from an ability to identify amongst the treated those who do, and those who do not, achieve the learning outcomes of training; and our investigation of the challenge of endogenous selection following treatment start that incorporation of this information presents. We present findings from estimation of training impacts for 25 to 55 year olds, associated with [1] *Preparation for Work at Level 1 or Below* (learning at or below European Qualification Framework, EQF *Level 2*); [2] *Level 1/Level 2 Math and/or English* (Maths and English at EQF *Level 2* or *Level 3*) and [3] *Full Level 2 and above* (more substantial learning aims, most of which are equivalent to EQF *Level 3*). For each of these categories of learning, we present four pairwise comparisons: [A] *all starters* and *currently*

²⁹ We have also experimented with smaller bins than the ones used in the main analysis, but this led to the loss of too many treated units (on average over 20% of the treated population).

untreated; [B] *fully treated* and *currently untreated*; [C] *partially treated* and *currently untreated*; and [D] *fully treated* and *partially treated*. The *fully treated* and *partially treated* are defined as such according to whether the educational administrative dataset records that they do, or do not, achieve learning outcomes, having enrolled for training.

We identify economically and statistically significant effects from training for the unemployed aged 25 to 55, which is predominantly initiated to remediate basic skills deficits that are likely to have persisted since the age of 16 for many. We observe employment premiums of between 4 and 7 ppts for *all starters* over the *currently untreated* aged 25 to 55 years old, with these impacts accruing from a point around 6 months from initiation for less substantial training commitments; and from a point 12 months from initiation for those that require a more substantial [hours] commitment. These are relatively large impacts, given employment rates of around 35 per cent for the group of unemployed being studied, who we may consider as hard to place in employment.

We can draw on existing evidence (Heckman and Smith, 1999; Lechner and Wunsch, 2013; and Caliendo et al. 2017) to invoke the CIA for comparison of *all starters* and *currently untreated*. However, there is currently no evidence to support our assumption that any unobserved factors driving selection into *full/partial* treatment following treatment start, are uncorrelated with potential outcomes, having conditioned on observed learning and employment histories prior to treatment start. Our investigation using distance to nearest FE training provider to instrument for possible endogenous selection into partial and full treatment following treatment start, attempts to overcome this deficit. Our findings suggest that the potential for such endogenous selection apparent in existing studies (Kluve et al., 2012; Fitzenberger et al., 2015), does not invalidate our invocation of the CIA for comparisons using the partially and fully treated.

Given this support for our identification strategy, we uncover a pattern of impacts that suggest the *partially* treated secure some form of return to human capital accumulation, and this is on average at a lower level than that achieved by the *fully* treated. These findings uncover a hierarchy of impacts consistent with those found in existing studies (for instance, Bell et al., 1995; Heckman et al., 1998a; Paul, 2015; Choe et al., 2015) and our estimates suggest that the employment premium secured by the *fully* treated over the *partially* treated is approximately 2 ppts in the period up to three years from training initiation.

These findings imply that an 'intention to treat' approach to the estimation of impacts, would not radically alter our conclusions regarding the overall efficacy of training interventions considered in this study. Our identification of an effect from *partial* treatment that is close to that for the *fully* treated, is reassuring in this sense when considering the approaches that dominate in the ALMP literatures. In contrast, evaluation approaches that attempt to capture returns to vocational learning using the partially treated to estimate counterfactual outcomes, risk understating effects.

However, whilst ppt treatment effects for the *fully* and *partially* treated are of similar magnitudes during the period up to three years from training initiation, estimated effects from *partial* treatment seem to exhibit less persistence, when compared to those accruing to the *fully* treated. For each of our categories of learning, the decline in estimated effects for *all*

starters which tends to appear from a point three years from training initiation, seems driven by the lesser persistence of effects accruing to the *partially* treated when compared to those of the *fully* treated.

Overall, we identify an important role for training at levels of learning between EQF Level 1 and Level 3, for unemployed individuals aged 25 to 55 who in most cases are attempting to remediate poor levels of attainment that have persisted since the age of 16. The specific costs associated with programmes such as Preparation for Life and Work; the Entry to Employment (E2E) pre-apprenticeship offer; more general aims at Level 1 and Below (including Maths/English) and also Level 2 are dependent on the number of FE contact hours (rather than the specific level) and the 'programme weighting'. The costs associated with these two components change over time for learners aged 16 and over³⁰. For example, a four-hour entry-level course in Preparation for Life and Work would be £24. However, the programmes we are capturing here are often of a longer duration, with for instance Level 1 functional [Maths/English] skills having a cost estimate equal to £724. One may consider that costs in the majority of courses being evaluated here, range from around £500 to £1,000. Given the impacts identified, it is likely that there is an overall net benefit to the exchequer³¹ and the short duration of many interventions; very low employment probabilities in the absence of intervention; and voluntary nature of the training considered, all work to limit lock-in effects.

Whilst we must be careful in suggesting that our study removes all concerns over endogenous selection into partial/full treatment following treatment start, it does suggest that future studies might usefully incorporate such information where it exists. This will allow additional consideration of the extent to which effects from partial treatment seem to exhibit a lesser persistence than those for the fully treated; and to shed further light on the differing magnitude of such effects. As McCall et al. (2016) make clear, the question of heterogeneity of treatment effects is an important topic for the literature. The data used in this study allow us to go further than many in attempting to identify such heterogeneity by age group, training type, level and treatment status. However, in doing so the study shows clearly the tension between this pursuit of treatment impacts for more specific groups and the appropriateness of existing methods.

³⁰ For specific examples of how the formula operates for 16/17 year olds see <u>https://dera.ioe.ac.uk/25358/1/Funding_rates_and_formula_2016_to_2017.pdf</u>

³¹ These figures do not consider the marginal social cost of public funds (MCF). For a more detailed discussion, see Greenberg, et al. (2011); Belfield, Crawford and Sibieta (2016).

References

Arni, P., Lalive, R. and van Ours, J.C. 2013. How effective are unemployment benefit sanctions: Looking beyond unemployment exit. *Journal of Applied Econometrics* 28: 1153–1178.

Barnow, B.S. and Smith, J. 2015. Employment and training programs. NBER Working Paper No. 21659.

Behrman, J., Cheng, Y. and Todd, P. 2004. Evaluating preschool programs when length of exposure to the program varies: a nonparametric approach. *The Review of Economics and Statistics* 86(1): 108-132.

Belfield, C., Crawford, C. and Sibieta, L. 2018. Long-run comparisons of spending per pupil across different stages of education. The Institute for Fiscal Studies, ISBN: 978-1-912805-04-4.

Bell, S., Orr, L., Blomquist, J and Cain, G. 1995. Program applicants as a comparison group in evaluating training programs: theory and a test. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bibby, D., Buscha, F., Cerqua, A., Thomson, D. and Urwin, P. 2014. Estimation of the labour market returns to qualifications gained in English Further Education. Department for Business, Innovation and Skills, Research Paper No. 195.

Bibby, D., Buscha, F., Cerqua, A., Thomson, D. and Urwin, P. 2015a. Impact of skills and training interventions on the unemployed: Phase I report. Department for Business, Innovation and Skills, Research Paper No. 158.

Bibby, D., Buscha, F., Cerqua, A., Thomson, D. and Urwin, P. 2015b. Impact of skills and training interventions on the unemployed: Phase II report. Department for Business, Innovation and Skills, Research Paper No. 158.

Biewen, M., Fitzenberger, B., Osikominu, A. and Paul, M. 2014. The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices. *Journal of Labor Economics* 32(4): 837-897.

Blackwell, M., Iacus, S. M., King, G. and Porro, G. 2009. CEM: Coarsened Exact Matching in Stata. *The Stata Journal* 9(4): 524-546.

Blundell, R., Costa Dias, M., Meghir, C. and van Reenen, J. 2004. Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association* 2: 569–606.

Blundell R., Dearden L. and Sianesi B. 2005. Evaluating the effect of education on earnings: Models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society Series A*, 168(3): 473-512.

Boockmann B., Thomsen S.L. and Walter T. 2014. Intensifying the use of benefit sanctions: an effective tool to increase employment? *IZA Journal of Labor Policy* 3(21): 1-19.

Buscha, F. and Urwin, P. 2013. Estimating the labour market returns to qualifications gained in English Further Education using the Individualised Learner Record (ILR). Department for Business, Innovation and Skills, final report.

Caliendo, M., Mahlstedt, R. and Mitnik, O.A. 2017. Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies. *Labour Economics* 46: 14–25.

Cameron, S.V. and Taber, C. 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112(1): 132–182.

Card, D. 1995. Using geographic variation in college proximity to estimate the return to schooling. In Aspects of labour market behaviour: Essays in honour of John Vanderkamp, ed. Louis N.Christofides,Kenneth E. Grant, and Robert Swidinsky, 201–222. Toronto: University of Toronto Press.

Card, D., Kluve, J. and Weber, A. 2010. Active labour market policy evaluations: A metaanalysis. *The Economic Journal* 120: 452-477.

Card, D., Kluve, J. and Weber, A. 2018. What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association* 16(3): 894–931.

Carneiro, P., Heckman, J.J. and Vytlacil, E.J. 2011. Estimating marginal returns to education. *American Economic Review* 101: 2754–2781

Choe, C., Flores-Lagunes, A. and Lee, S.J. 2015. Do dropouts with longer training exposure benefit from training programs? Korean evidence employing methods for continuous treatments. *Empirical Economics* 48: 849-881.

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. 2016. From Late to MTE: Alternative Methods for the Evaluation of Policy Interventions, *IZA Discussion Paper* No. 10056.

De Crombrugghe, D., Espinoza, H. and Heijke, H. 2010. Determinants of dropout behavior in a job training programme for disadvantaged youth. Working paper ROA-RM-2010/8, Research Centre for Education and the Labor Market, Maastricht University.

Dickson, M. 2009. The causal effect of education on wages revisited. IZA Discussion Paper No 4419.

Dolton, P. and Smith, J. 2011. The impact of the UK New Deal for Lone Parents on benefit receipt. IZA Discussion Paper No. 5491.

Dorsett, R. 2006. The New Deal for Young People: Effect on the labour market status of young men. *Labour Economics* 13: 405–422.

Dorsett, R., Gray, H., Speckesser, S. and Stokes, L. 2019. Estimating the impact of Traineeships. Department for Education, final report.

Fay, R.G. 1996. Enhancing the effectiveness of active labour market policies: Evidence from programme evaluations in OECD countries. Labour Market and Social Policy Occasional Paper 18, OECD, Paris.

Filges, T., Smedslund, G. and Jørgensen, A.-M.K. 2016. Active labour market programme participation for unemployment insurance recipients: A systematic review". *Research on Social Work Practice*: 1-29.

Fitzenberger, B., Osikominu, A. and Völter, R. 2006. Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany. IZA Discussion Paper No. 2121.

Fitzenberger, B., Osikominu, A., Paul, M. 2015. The effects of training incidence and duration on labor market transitions. Working Paper (August 2015 version).

Fredriksson, P. and Johansson, P. 2008. Dynamic treatment assignment. *Journal of Business and Economic Statistics* 26(4): 435-445.

Frenette, M. 2004. Access to college and university: Does distance to school matter?. *Canadian Public Policy*, 20(4), 427–443.

Frölich, M. 2004. Programme evaluation with multiple treatments. *Journal of Economic Surveys* 18: 181–224.

Greenberg, D., Knight, G., Speckesser, S. and Hevenstone, D. 2011. Improving DWP assessment of the relative costs and benefits of employment programmes. Department for Work and Pensions Working Paper No 100.

Hedges, S., Patrignani, P. and Conlon, G. 2018. Settling the counterfactual debate: Is there a preferable counterfactual when estimating the returns to vocational qualifications? CVER Research Discussion Paper No 13.

Heckman, J., Smith, J. and Taber, C. 1998a. Accounting for dropouts in evaluations of social programs. *The Review of Economics and Statistics* 80(1): 1–14.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. 1998b. Characterizing selection bias using experimental data. *Econometrica* 66(5): 1017–1098.

Heckman, J. and Smith, J. 1999. The pre-program earnings dip and the determinants of participation in a social program: Implications for simple program evaluation strategies. *The Economic Journal* 109: 313–348.

Heckman, J., Hohmann, N., Smith, J. and Khoo, M. 2000. Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics*

115(2): 651–694.

Hirshleifer, S., McKenzie, D., Almeida, R. and Ridao-Cano, C. 2015. The impact of vocational training for the unemployed: Experimental evidence from Turkey. *The Economic Journal* 126(597): 2115–2146.

Huber, M., Lechner, M. and Wunsch, C. 2013. The performance of estimators based on the propensity score. *Journal of Econometrics* 175: 1-21.

lacus, S. M., King, G. and Porro, G. 2011. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association* 106(493): 345-361.

Iacus, S. M., King, G. and Porro, G. 2012. Causal inference without balance checking: Coarsened exact matching. *Political Analysis* 20, 1-24.

Imbens, G. W. and Angrist, J. D. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62(3): 467–475.

Jepsen, C., Troske, K. and Coomes, P. 2014. The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics* 32(1): 95-121.

Kjellström, C. and Regner, H. 1999. The effects of geographical distance on the decision to enrol in university education. *Scandinavian Journal of Educational Research* 43(4): 335–348.

Kluve, J. 2010. The effectiveness of European active labor market programs. *Labour Economics* 17: 904-918.

Kluve, J., Rinne, U., Uhlendorff, A. and Zhao, Z. 2012. Endogenous exit bias in training programs for unemployed workers. IZA Discussion Paper No. 6565.

Lechner, M., Miquel, R. and Wunsch, C. 2007. The curse and blessing of training the unemployed in a changing economy: The case of East Germany after unification. *German Economic Review* 8: 468–507.

Lechner, M., Miquel, R. and Wunsch, C. 2011. Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association* 9(4): 742-784.

Lechner, M. and Wunsch, C. 2013. Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Economics* 21: 111-121.

McCall, B., Smith, J.A. and Wunsch, C. 2016. Government-sponsored vocational education for adults. Handbook of the Economics of Education 5: 479-652.

OECD. 2014. Connecting people with jobs: Activation policies in the United Kingdom. OECD Publishing.

Patrignani, P. and Conlon, G. 2011. The long-term effect of vocational qualifications on labour market outcomes. Department for Business, Innovation and Skills, Research Paper No. 47.

Paul, M. 2015. Many dropouts? Never mind! - Employment prospects of dropouts from training programs. *Annals of Economics and Statistics* 119-120: 235-267.

Plesca, M. and Smith, J. 2007. Evaluating multi-treatment programs: Theory and evidence from the U.S. job training partnership act. *Empirical Economics* 32(2-3): 491-528.

Rubin, D.B. 1986. Comment: Which ifs have causal answers. *Journal of the American Statistical Association* 81: 961–962.

Sianesi, B. 2004. An evaluation of the Swedish system of active labour market programs in the 1990s. *The Review of Economics and Statistics* 1: 133–155.

Sianesi, B. 2008. Differential effects of active labour market programs for the unemployed. *Labour Economics* 15: 370–399

Spiess, K. and Wrohlich, K. 2010. Does distance determine who attends a university in Germany? *Economics of Education Review* 29: 470–479

Stuart, E.A. 2010. Matching methods for causal inference: a review and a look forward. *Statistical Science* 25(1): 1–21.

Stuart, E.A. and Rubin, D.B. 2007. Best practices in quasi-experimental designs: Matching methods for causal inference. Chapter 11 (pp. 155-176) in *Best practices in quantitative social science*, J. Osborne (Ed.): Thousand Oaks, CA: Sage Publications.

Van den Berg, G.J., Bozio, A. and Costa Dias, M. 2014. Policy discontinuity and duration outcomes. IZA Discussion Papers No. 8450.

Walters, C. 2018. The Demand for Effective Charter Schools. *Journal of Political Economy*, 126(6): 2179-2223.

Wunsch, C. 2016. How to minimize lock-in effects of programs for unemployed workers", *IZA World of Labor 288*.

	Percentage	point employ	ment in years	after learning s	spell start	_
		1 st Year	2 nd Year	3 rd Year	4 th Year	
	All Starters & Currently Untreated	-0.012*** (0.004)	-0.002 (0.004)	-0.002 (0.004)	-0.003 (0.004)	
1 for Work	Fully Treated & Currently Untreated	-0.007 (0.005)	0.011** (0.005)	0.014** (0.006)	0.012** (0.006)	
Preparation	Partially Treated & Currently Untreated	-0.024*** (0.005)	-0.022*** (0.006)	-0.025*** (0.006)	-0.024*** (0.006)	
	Fully Treated & Partially Treated	0.026*** (0.006)	0.038*** (0.007)	0.040*** (0.007)	0.035*** (0.007)	_
racy	All Starters & Currently Untreated	0.005 (0.005)	0.014*** (0.005)	0.017*** (0.005)	0.008* (0.005)	
.2 Literacy and/or Numera	Fully Treated & Currently Untreated	0.010* (0.006)	0.031*** (0.007)	0.034*** (0.007)	0.026*** (0.007)	
	Partially Treated & Currently Untreated	-0.003 (0.007)	-0.011 (0.008)	-0.011 (0.008)	-0.018** (0.008)	
Г 1 7	Fully Treated & Partially Treated	0.020*** (0.006)	0.037*** (0.007)	0.036*** (0.007)	0.034*** (0.007)	_
	All Starters & Currently Untreated	0.001 (0.004)	0.042*** (0.005)	0.047*** (0.005)	0.046*** (0.005)	
and above	Fully Treated & Currently Untreated	-0.003 (0.006)	0.064*** (0.007)	0.075*** (0.007)	0.073*** (0.007)	
Full Level 2 ¿	Partially Treated & Currently Untreated	0.002 (0.006)	0.022*** (0.007)	0.022*** (0.007)	0.023*** (0.007)	
	Fully Treated & Partially Treated	0.008 (0.008)	0.039*** (0.009)	0.048*** (0.009)	0.051*** (0.009)	Note: In th 'Preparation for Worl analysis, - there ar

Appendix Table A1: Summary of employment outcomes: aged 18 to 24

12,974 treatment starters (7,279 fully treated and 5,695 partially treated) and the 'currently untreated' group includes 440,833 individuals who do not initiate treatment during the period of analysis. In the 'Level 1/Level 2 Math and/or English' analysis, there are 15,702 treatment *starters* (9,207 *fully* treated and 6,495 *partially* treated) and the 'currently untreated' group includes 437,211 individuals who do not initiate treatment during the period of analysis. In the 'Full Level 2 and above' analysis, there are 9,342 treatment *starters* (4,604 *fully* treated and 4,738 *partially* treated) and the 'currently untreated' group includes 445,006 individuals who do not initiate treatment during the period of analysis. ***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.

		Fully Treated	Partially Treated		Currently Untreated	
	Variables	Treated		Difference		Difference
	Proportion of Males	0.618	0.618	(0)	0.618	(0)
	Age at start	40.57	40.61	(-0.04)	40.63	(-0.06)
	White ethnicity	0.739	0.739	(0)	0.739	(0)
	Months in employment in the last 18 months	7.42	7.38	(0.04)	7.41	(0.01)
	Disability dummy	0.324	0.324	(0)	0.324	(0)
/ork	Offender dummy	0.008	0.009	(-0.001)	0.006	(0.002)**
N N	Ever lone parent dummy	0.141	0.153	(-0.012)**	0.148	(-0.007)
n fo	Children dummy	0.199	0.207	(-0.008)*	0.210	(-0.011)**
ratic	Ever asylum seeker dummy	0.005	0.006	(-0.001)	0.006	(-0.001)
epai	Prior LMS referrals	0.200	0.199	(0.001)	0.196	(0.004)
Ъл	Prior mandatory referrals	0.206	0.211	(-0.005)	0.201	(0.005)
	Prior ILR started	1.26	1.18	(0.08)**	0.79	(0.47)***
	Prior ILR achieved	0.83	0.75	(0.08)*	0.49	(0.34)***
	Unemployment rate	5.52	5.50	(0.02)	5.56	(-0.04)
	IMD score	25.10	24.77	(0.33)***	25.01	(0.09)
	Employment ID score	19,299	18,663	(636)***	18,988	(311)
	Proportion of Males	0.738	0.738	(0)	0.738	(0)
	Age at start	37.29	37.26	(0.03)	37.41	(-0.12)
	White ethnicity	0.710	0.710	(0)	0.710	(0)
	Months in employment in the last 18 months	7.01	6.87	(0.14)*	6.88	(0.13)**
lish	Disability dummy	0.355	0.355	(0)	0.355	(0)
Eng	Offender dummy	0.010	0.011	(-0.001)	0.008	(0.002)***
lor I	Ever lone parent dummy	0.164	0.172	(-0.008)*	0.159	(0.005)
and	Children dummy	0.255	0.257	(-0.002)	0.246	(0.009)**
ths	Ever asylum seeker dummy	0.008	0.008	(0)	0.008	(0)
Mai	Prior LMS referrals	0.268	0.281	(-0.013)	0.270	(-0.002)
11/2	Prior mandatory referrals	0.280	0.287	(-0.007)	0.283	(-0.003)
Ľ	Prior ILR started	1.44	1.41	(0.03)	0.90	(0.54)***
	Prior ILR achieved	0.87	0.81	(0.06)*	0.55	(0.32)***
	Unemployment rate	5.56	5.58	(-0.02)	5.59	(-0.03)
	IMD score	25.90	25.79	(-0.01)	25.45	(0.45)**
	Employment ID score	20,007	20,199	(-0.192)	19,397	(610)***
Leve	Proportion of Males	0.697	0.697	(0)	0.697	(0)
nll I	Age at start	36.33	36.26	(0.07)	36.56	(-0.23)
ц	White ethnicity	0.672	0.672	(0)	0.672	(0)
	Months in employment in the last 18 months	7.67	7.64	(0.03)	7.65	(0.02)
	Disability dummy	0.229	0.229	(0)	0.229	(0)
	Offender dummy	0.005	0.005	(0)	0.005	(0)
	Ever lone parent dummy	0.141	0.154	(-0.013)**	0.149	(-0.008)

Table A2: Average pre-treatment differences between Fully Treated, PartiallyTreated and Currently Untreated in the matched sample

Children dummy	0.244	0.240	(0.004)	0.230	(0.014)*
Ever asylum seeker dummy	0.008	0.007	(0.001)	0.008	(0)
Prior LMS referrals	0.214	0.210	(0.004)	0.209	(0.005)
Prior mandatory referrals	0.219	0.219	(0)	0.232	(-0.013)**
Prior ILR started	1.61	1.63	(-0.02)	0.84	(0.77)***
Prior ILR achieved	0.98	0.91	(0.07)	0.70	(0.28)***
Unemployment rate	5.70	5.61	(0.09)	5.72	(-0.02)
IMD score	26.01	25.50	(0.51)*	25.59	(0.42)
Employment ID score	20,911	19,888	(1,023)**	19,670	(1,241)***

Note: Disability indicator captures at least one of the following physical issues: Mobility, Manual Dexterity, Physical Coordination, Continence, "Ability to Lift, carry or otherwise move everyday objects", "Speech, hearing or eyesight", "Memory or ability to concentrate, learn or understand", Perception of the risk of physical danger, Severe Disfigurement. ***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.

Table A3: Summary of employment outcomes using PSM after CEM

	Percentage	point employr	ment in years a	fter learning s	pell start
		1 st Year	2 nd Year	3 rd Year	4 th Year
	All Starters & Currently Untreated	0.045*** (0.003)	0.047*** (0.003)	0.049*** (0.003)	0.042*** (0.003)
for Work	Fully Treated & Currently Untreated	0.047*** (0.003)	0.050*** (0.003)	0.053*** (0.004)	0.049*** (0.004)
Preparation	Partially Treated & Currently Untreated	0.039*** (0.003)	0.038*** (0.004)	0.035*** (0.004)	0.030*** (0.004)
	Fully Treated & Partially Treated	0.012** (0.006)	0.013** (0.006)	0.017*** (0.006)	0.017*** (0.007)
acy	All Starters & Currently Untreated	0.033*** (0.002)	0.038*** (0.003)	0.032*** (0.003)	0.029*** (0.003)
d/or Numer	Fully Treated & Currently Untreated	0.033*** (0.003)	0.041*** (0.003)	0.040*** (0.003)	0.044*** (0.004)
/L2 Literacy an	Partially Treated & Currently Untreated	0.031*** (0.003)	0.026*** (0.004)	0.013*** (0.004)	0.004 (0.004)
L1	Fully Treated & Partially Treated	0.007* (0.004)	0.017*** (0.005)	0.020*** (0.005)	0.036*** (0.005)
	All Starters & Currently Untreated	0.025*** (0.003)	0.059*** (0.003)	0.068*** (0.003)	0.065*** (0.004)

Fully Treated &	0.014***	0.062***	0.075***	0.075***	
Currently	(0.003)	(0.004)	(0.004)	(0.004)	
Untreated					
Partially	በ በ37***	0 056***	0 062***	0 05/***	
	0.007	0.000	0.002	0.034	
I reated &	(0.004)	(0.004)	(0.004)	(0.005)	
Currently					
Untreated					
					Note: We
Fully Treated &	-0.008	0.015**	0.013*	0.019***	used K-
Partially	(0.005)	(0.006)	(0.007)	(0.007)	nearest
Treated	、 ,		、	、 ,	neighbour
					matching with

K=10 and a caliper=0.002. We used CEM before PSM in order to drop the observations falling off the common support. ***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.

Table A4: Summary of employment outcomes using IPW after CEM

	Percentage point employment in years after learning spell start							
		1 st Year	2 nd Year	3 rd Year	4 th Year			
	All Starters & Currently Untreated	0.039*** (0.003)	0.046*** (0.003)	0.044*** (0.003)	0.042*** (0.003)			
for Work	Fully Treated & Currently Untreated	0.041*** (0.003)	0.048*** (0.003)	0.050*** (0.003)	0.047*** (0.004)			
Preparation	Partially Treated & Currently Untreated	0.036*** (0.003)	0.036*** (0.003)	0.033*** (0.004)	0.029*** (0.004)			
	Fully Treated & Partially Treated	0.009* (0.005)	0.010** (0.005)	0.020*** (0.006)	0.021*** (0.006)			
racy	All Starters & Currently Untreated	0.034*** (0.002)	0.040*** (0.002)	0.037*** (0.002)	0.032*** (0.002)			
d/or Nume	Fully Treated & Currently Untreated	0.035*** (0.002)	0.046*** (0.003)	0.042*** (0.003)	0.044*** (0.003)			
IL2 Literacy an	Partially Treated & Currently Untreated	0.030*** (0.003)	0.027*** (0.003)	0.015*** (0.003)	0.002 (0.003)			
L1,	Fully Treated & Partially Treated	0.006 (0.004)	0.016*** (0.004)	0.024*** (0.004)	0.040*** (0.004)			
	All Starters & Currently	0.028*** (0.002)	0.058*** (0.003)	0.063*** (0.003)	0.059*** (0.003)			

Untreated					
Fully Treated &	0.013***	0.061***	0.068***	0.071***	
Currently Untreated	(0.003)	(0.003)	(0.004)	(0.004)	
Partially	0.036***	0.057***	0.060***	0.053***	
Treated & Currently	(0.004)	(0.004)	(0.004)	(0.005)	
Untreated					Note: We used CEM
Fully Treated &	-0.012**	0.009	0.010*	0.015**	before IPW in
Partially Treated	(0.005)	(0.006)	(0.006)	(0.006)	order to drop
					observations

falling off the common support.

***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.

Table A5: Summary of sustained employment outcomes

	Percentage point sustained employment in years after learning spell start					
		1 st Year	2 nd Year	3 rd Year	4 th Year	
	All Starters & Currently Untreated	0.039*** (0.003)	0.048*** (0.003)	0.045*** (0.003)	0.040*** (0.003)	
for Work	Fully Treated & Currently Untreated	0.040*** (0.003)	0.051*** (0.004)	0.049*** (0.004)	0.046*** (0.004)	
Preparation	Partially Treated & Currently Untreated	0.037*** (0.004)	0.038*** (0.004)	0.032*** (0.004)	0.026*** (0.004)	
	Fully Treated & Partially Treated	0.009* (0.005)	0.013** (0.006)	0.010* (0.006)	0.010 (0.007)	
acy	All Starters & Currently Untreated	0.027*** (0.003)	0.034*** (0.003)	0.026*** (0.003)	0.022*** (0.003)	
d/or Numer	Fully Treated & Currently Untreated	0.027*** (0.003)	0.037*** (0.003)	0.033*** (0.003)	0.034*** (0.004)	
lL2 Literacy an	Partially Treated & Currently Untreated	0.024*** (0.004)	0.029*** (0.004)	0.012*** (0.004)	0.006 (0.004)	
ΓΤ	Fully Treated & Partially Treated	0.016*** (0.004)	0.018*** (0.005)	0.022*** (0.005)	0.030*** (0.005)	

All Starters & Currently Untreated	0.027*** (0.003)	0.058*** (0.003)	0.063*** (0.003)	0.060*** (0.003)	
Fully Treated & Currently Untreated	0.015*** (0.003)	0.060*** (0.003)	0.071*** (0.004)	0.069*** (0.004)	
Partially Treated & Currently Untreated	0.043*** (0.004)	0.057*** (0.004)	0.056*** (0.004)	0.048*** (0.005)	Note: ***, **, * denote significance at the 1-, 5-,
 Fully Treated & Partially Treated	-0.012** (0.005)	0.016*** (0.005)	0.017*** (0.005)	0.030*** (0.006)	and 10- percent level, respectively.

	Percentage point employment in years after learning spell start					
		1 st Year	2 nd Year	3 rd Year	4 th Year	_
	All Starters & Currently Untreated	0.044*** (0.003)	0.053*** (0.004)	0.052*** (0.004)	0.045*** (0.004)	
Preparation for Work	Fully Treated & Currently Untreated	0.047*** (0.004)	0.057*** (0.004)	0.058*** (0.004)	0.050*** (0.005)	
	Partially Treated & Currently Untreated	0.046*** (0.003)	0.039*** (0.004)	0.033*** (0.004)	0.027*** (0.004)	
	Fully Treated & Partially Treated	0.006 (0.005)	0.017*** (0.006)	0.023*** (0.006)	0.022** (0.007)	_
racy	All Starters & Currently Untreated	0.032*** (0.002)	0.038*** (0.003)	0.032*** (0.003)	0.028*** (0.003)	
-2 Literacy and/or Nume	Fully Treated & Currently Untreated	0.034*** (0.003)	0.042*** (0.003)	0.040*** (0.004)	0.038*** (0.004)	
	Partially Treated & Currently Untreated	0.028*** (0.004)	0.024*** (0.004)	0.019*** (0.004)	0.007* (0.004)	
ГJ	Fully Treated & Partially Treated	0.011*** (0.004)	0.021*** (0.004)	0.019*** (0.004)	0.029*** (0.005)	_
	All Starters & Currently Untreated	0.030*** (0.003)	0.055*** (0.003)	0.062*** (0.003)	0.063*** (0.003)	
snj	Fully Treated & Currently Untreated	0.018*** (0.003)	0.060*** (0.003)	0.074*** (0.003)	0.071*** (0.004)	
FL2 F	Partially Treated & Currently Untreated	0.035*** (0.004)	0.053*** (0.004)	0.053*** (0.005)	0.054*** (0.005)	
	Fully Treated & Partially Treated	-0.006 (0.005)	0.012** (0.006)	0.017*** (0.006)	0.016*** (0.006)	Note: ***, denote significanc at the 1-, 5

Table A6: Summary of employment outcomes when only considering the sub-intervals in the first 12 of the 18 months prior to ""mandatory" New Deal services

percent level, respectively.

	Percentage point employment in years after learning spell start					_
		1 st Year	2 nd Year	3 rd Year	4 th Year	_
	All Starters & Currently Untreated	0.044*** (0.002)	0.053*** (0.002)	0.051*** (0.002)	0.044*** (0.003)	
for Work	Fully Treated & Currently Untreated	0.044*** (0.002)	0.056*** (0.003)	0.055*** (0.003)	0.052*** (0.003)	
Preparation	Partially Treated & Currently Untreated	0.043*** (0.003)	0.043*** (0.003)	0.037*** (0.003)	0.030*** (0.003)	
	Fully Treated & Partially Treated	0.012*** (0.004)	0.021*** (0.005)	0.018*** (0.005)	0.017*** (0.005)	_
racy	All Starters & Currently Untreated	0.032*** (0.002)	0.038*** (0.002)	0.031*** (0.003)	0.027*** (0.003)	
nd/or Nume	Fully Treated & Currently Untreated	0.032*** (0.002)	0.040*** (0.002)	0.038*** (0.003)	0.040*** (0.003)	
lL2 Literacy aı	Partially Treated & Currently Untreated	0.029*** (0.003)	0.022*** (0.003)	0.014*** (0.003)	0.001 (0.003)	
Г Т	Fully Treated & Partially Treated	0.015*** (0.003)	0.024*** (0.003)	0.020*** (0.004)	0.033*** (0.004)	_
	All Starters & Currently Untreated	0.031*** (0.002)	0.064*** (0.002)	0.072*** (0.003)	0.074*** (0.003)	
lus	Fully Treated & Currently Untreated	0.020*** (0.003)	0.069*** (0.003)	0.080*** (0.003)	0.082*** (0.003)	
FL2 p	Partially Treated & Currently Untreated	0.049*** (0.003)	0.061*** (0.003)	0.063*** (0.003)	0.057*** (0.004)	
	Fully Treated & Partially Treated	-0.006 (0.004)	0.013*** (0.004)	0.024*** (0.004)	0.028*** (0.005)	Note: Matched (using CEM) on number of _ months in

Table A7: Summary of employment outcomes using larger bins

employment between month (t-1) and month(t-60) split into tertiles; number of months on active benefits between month (t-1) and month (t-60) split at the median; number of months between the beginning of the unemployment spell and the beginning of training split at the median; caseworker referral, gender, age group split at the median, white/non-white ethnicity, and number of prior ILR aims started, split at the median.

***, **, * denote significance at the 1-, 5-, and 10-percent level, respectively.