Does Literacy and Numeracy learning help the unemployed find a job? Evidence from England using ILR-WPLS admin data

David Bibby, Augusto Cerqua, Dave Thomson and Peter Urwin

June 2015

Presentation SOLE-EALE 26th June

Work commissioned by the UK Department for Business, Innovation and Skills, and Department for Work and Pensions

EMBARGOED FINDINGS, PLEASE DO NOT QUOTE WITHOUT AUTHORS' PERMISSION

Please contact: Prof. Peter Urwin
Centre for Employment Research
University of Westminster
35 Marylebone Road, London NW1 5LS
020 7911 5000
email: urwinp@westminster.ac.uk

Abstract: In this paper we use Coarsened Exact Matching (CEM) to estimate the returns to achievement in early literacy and numeracy interventions for a population of individuals in England who have an unemployment claim start date between April 2006 and 2008. The analysis makes use of the ILR-WPLS-LMS-ND administrative dataset, which contains information on learning, details of any benefit claims, earnings information and employment, amongst others. One estimate of value added is obtained by comparing (i) the employment returns of unemployed individuals who have a Further Education (FE is UK equivalent of US Community Colleges) literacy/numeracy learning aim that they achieve; with (ii) the returns of matched individuals who have the same FE learning aim, but drop-out and do not achieve. Another estimate is obtained by comparing (i) the returns to unemployed individuals who have an FE literacy/numeracy learning aim that they achieve; with (iii) the returns of a matched group of individuals who have no learning aim identified in the FE learner dataset. This is one of the few times a study has been able to differentiate the returns to training, according to whether the unemployed individual achieves the learning outcomes of the course; and our choice of population allows the tracking of returns 60 months on from claim start date and matching on up to 8 years of prior labour market information. We identify statistically significant returns to basic literacy/numeracy interventions in the months after claim start date, suggesting that returns may be missed in many studies, because of (i) an inability to differentiate achievers and non-achievers and (ii) the inability to create valid counterfactual estimates that capture negative selection into low level vocationally-oriented learning.

Key words: Short-term unemployment, literacy interventions, employment returns

JEL Classification: I21, I26, I38, J64

We thank colleagues at BIS, particularly Karen Woolgar (nee Grierson), Adrian Jones, Peter Vallely and Anthony Harris; and colleagues at DWP, particularly Richard Ward.

1. Introduction

In this paper we consider an initial population of 2.3 million individuals in England with an unemployment benefit claim start date between 6th April 2006 and 5th April 2008[1]. We match administrative information from benefit and employment records (the Work and Pensions Longitudinal Study, WPLS), to information held on the Labour Market System (LMS), which is used by advisors within English Jobcentre Plus offices to record the detail of individual referrals to a variety of short interventions. We are interested in the impact of training carried out earlier in an unemployment spell, and during this phase the majority of LMS referrals will be to brief sessions on CV writing, interview technique and other basic support and guidance. In the analysis presented here, we do not pursue the detail contained within the matched LMS data, but simply split the initial population into 1.52m (65.5%) who do not have any form of referral 'flagged' within the LMS data, and 0.73m (31.5%) who have at least one such referral to basic support and guidance flagged by a Jobcentre Plus advisor[2].

One of the main contributions of our paper is the ability to then match these data to administrative (Individualised Learner Record, ILR) information on all registered learning aims at an English Further Education (FE) Institution between the 2002/2003 and 2012/2013 academic years (thus creating the ILR-WPLS-LMS-ND: see Bibby et al., 2012). English FE Institutions are broadly equivalent to US Community Colleges, where *Certificates* and *Diplomas* are offered in predominantly technical (vocational) programs, and *Associate Degrees* are used as a possible route into Higher Education. UK government spending on FE has amounted to approximately £4bn per annum[3] in recent years and the number of adult learners (aged 19+) participating in government-funded FE was 3.28m in 2012/13. Much of the learning undertaken by those aged under 19 in England is also carried out in FE, for instance in the same year close to 200,000 under the age of 19 participated in funded apprenticeships[4]. Most importantly for this paper, between 2006 and 2008 the majority of government-funded literacy and numeracy programmes for the unemployed in this population will have been delivered within FE Institutions.

Our analysis focuses on the 121,346 individuals who have a Level 1 or Level 2 literacy and/or numeracy learning aim(s) (with 43,545 having only literacy aim; 32,563 only numeracy and 45,238 both), identified in the matched FE data[5]. We use a monotonic imbalance bounding (MIB) technique (see Iacus et al., 2011), Coarsened Exact Matching (CEM), to create estimates of the value added of these targeted literacy and numeracy interventions. Our focus of analysis is on the impact of this FE learning, with information on other interventions used in the process of matching.

---

[1] These are individuals on Jobseekers Allowance , who are mandated to engage in active jobsearch.

[2] The remaining 3% are dropped from the analysis, as we observe them being fast-tracked to more substantial Active Labour Market Programmes (ALMP) during this early period of unemployment. This can be because they have only just re-entered unemployment from a previous-'adjacent' spell; as a result of some specific characteristic seen as presenting a potential barrier to securing employment (such as being a Lone Parent) or in other ways that we may not be able to capture in the data.

[3] With the recent round of austerity cutting this from around £3.93 bn to £3.85 bn between 2011/2012 and 2012/2013.

[4] *Skills Funding Agency Statistical First Release*: SFA/SFR24, June 2014

[5] Level 1 literacy qualifications will tackle basic *functional* literacy and Level 2 is equivalent to that attempted by English secondary school pupils at the age of 15. Broadly equivalent to the US Community College *Certificates* and *Diplomas*.

Matching on a variety of characteristics, including more than 60 months of prior labour market and learning histories (8 years in many cases), we estimate employment returns in each month up to 60 months on from learning start date. One estimate of employment returns is created by comparing the outcomes of (i) those who *achieve* the L1/L2 literacy and/or numeracy aim identified in the ILR (FE) records, with those who (ii) have the same learning aim recorded, but drop-out without achieving the learning objectives. In addition, we create estimates of impact by comparing the returns of (i) *achievers*, with those of a more general population who (iii) have no recorded ILR FE learning aim within the initial 'Short-Term' period of Unemployment (STU) – matching exactly on the extent to which achievers and those with no 'in-scope' FE learning intervention have, or have not, been flagged as having some form of basic need (in the LMS) by Jobcentre Plus advisors.

The individual's initial claim start date is considered as time (t) equal to zero, and their expected date of referral to an ALMP ('X') is calculated from this[6]. For policy purposes, the individual will be considered as entering a period of Long Term Unemployment (LTU) if their spell persists beyond X. The figures on referrals and interventions (training or otherwise) described to this point are those identified as being 'in-scope', in that they occur between t=0 and X[7]. All outcomes are referenced to the learning start date and because our FE learning spells could be started by individuals when they are not necessarily on benefits, we limit our analysis to those who are observed on 'active' (i.e. job seeking) unemployment benefits at the point where learning starts (or the imputed start date of learning for the 'No FE' control group, following Lechner et. al. 2011).

Our study is clearly located within the ALMP literature (for instance, Dorsett, 2006; Card et al., 2010; Kluve, 2010; Andersen and Svarer, 2012; Biewen et al., 2014), but also a growing literature that considers the returns to vocational learning using administrative data from educational systems (Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Jepsen, Troske and Coomes, 2014; Bibby et. al. 2014). The majority of existing studies are unable to distinguish achievement and non-achievement (or drop-out) in programmes aimed at supporting the unemployed; they rarely have such lengthy periods over which to consider labour market histories and outcomes (though see Lechner et. al. 2011); few have such substantial numbers in both treatment and control groups; and in addition we are able to match on the detail of previous advisor referral to basic support and FE learning spells.

Section 2 describes the Data and Method, with Section 2.1 setting out the process of matching undertaken to create the ILR-WPLS dataset for our populations of unemployed individuals; and Section 2.2 describes selection of within-scope interventions and referrals for the populations of interest. Section 2.3 describes our approach to matching and Section 3 presents the preliminary findings. Section 4 concludes.

[6] X is 6 months for most 18 to 24 year-olds and 18 months for the majority of those aged 25+ in the period we consider. There will be some variation to this as other characteristics can be taken into account by Jobcentre Plus advisors, but this will impact a relatively small proportion and as already suggested we remove from consideration all those who are fast-tracked to more substantial ALMP interventions (i.e. their observed X and possibly therefore expected X are much earlier than in the majority of instances).

[7] In recognition of the potential margin for error around the expected claim start date of X, a 'fuzzy' X is created covering the period between X – 2 weeks and X + 2 weeks. 'Intervention' is an all-encompassing term, which reflects any type of referral in the LMS (training or otherwise) and any recorded ILR aim at an FE Institution.

2. Data and Method

This study focuses on unemployed individuals with a First or Only '*Active Benefits*'[8] claim start date between 6th April 2006 and 5th April 2008, who are resident in England and who may be attending an FE institution in England[9]. We consider the impact of training interventions undertaken during the initial Short-Term Unemployed (STU) period of unemployment. Differential consideration of the STU and the Long Term Unemployed (LTU) is driven by methodological considerations and also the policy context. Methodologically, we are likely to observe differences in estimated returns to training delivered early in an unemployment spell, compared to that delivered much later in a spell (for those who experience longer spells). Also to accommodate the policy context, we need to consider training undertaken by individuals prior to any referral to an ALMP (in this case the New Deal programme), separately to that delivered as part of the New Deal; as this is the point at which [predominantly] voluntary interventions become mandatory.

There is a distinct point in an individual's unemployment spell when we expect them to be referred to some form of ALMP intervention. For the purposes of policy, the individual moves from being considered as STU to LTU. This point in time varies according to the age of the individual, the specific policy regime and other relevant factors. In our general discussions we refer to the point where an individual is expected to become LTU, as the 'X month' of their unemployment duration. For instance, we expect an individual aged 18 to 24, with a claim start date falling within our inflow window, to be referred to the New Deal for Young People (NDYP)[10] at a point 6 months on from their claim start date – X will be equal to 6 months, whilst for those Aged 25+ this will be 18 months from claim start date.

We select individuals flowing on to benefits between April 2006 and 2008, so that we potentially have information on up to 8 years of prior labour market history information and five years of outcomes; with our focus on evaluation of literacy interventions delivered between claim start date and X, within Further Education settings.

2.1 Creation of the ILR-WPLS dataset

The ILR-WPLS dataset for this project consists of linked administrative data on:

- Learning in further education, sourced from the Individualised Learner Record (ILR);
- Receipt of state benefits, sourced from the National Benefits Database (NBD);
- Spells in employment, sourced from HMRC P45 records;
- Earnings, sourced from HMRC P14 records;
- Mandated or signposted referrals by Job Centre Plus staff to interventions for the unemployed, sourced from DWP Labour Market System (LMS) and New Deal (ND) evaluation databases.

[8] JobSeekers Allowance (JSA) mandates active job search and job availability.
[9] The ILR records training in all English FE Institutions, whilst the WPLS covers unemployed individuals resident in England, Scotland and Wales (not Northern Ireland). Clearly we could have individuals living in Wales/Scotland, close to the border with England, and attending an English FE – and vice versa. However, the numbers are likely to be relatively small and we therefore limit ourselves to the unemployed resident in England and training that takes place in English FE institutions. The population of England is approximately 86% of the population of England, Scotland and Wales.
[10] An ALMP aimed at the LTU aged 18 to 24.

The first step in compiling the data for analysis is the creation of a unique identifier linking an individual's records across the constituent databases and data files. Each constituent data source has its own unique individual identifier which may not be internally unique, and which (before transfer to the project team) may incorrectly ascribe records to the same individual; and conversely not match records correctly to the same individual. Additionally, the Department of Work and Pensions provided a table of record linkages derived from fuzzy matching of the constituent datasets. We use the 3 identifiers from the ILR, National Benefits Database (CCORCID) and the HMRC person-instance-ID (PID) to construct an over-arching Person-key to link records in the data sources reliably to the same individual.

Whereas data sourced from DWP or HMRC contain references that identify distinct individuals, allowing for the fact that some individuals have multiple identifiers, there are no such references in the ILR. Distinct ILR learners can only be identified on the basis of linkage to a DWP or HMRC dataset. For the purposes of this project then, all individuals have a CCORCID and may have no, one or many ILR learner records and no, one or many person-instance-ids. We use the three identifiers in a process of 'record chaining' to show the relationships between them, accepting that we have insufficient information to indubitably decide which linkages are correct and which are not. As a consequence, we develop a set of procedures to arbitrate between competing matching possibilities according to circumstances, but without rejecting the possibility of the same individual having more than one CCORCID or PID. These procedures lead to the creation of our Person-Key, which identifies the same individual in the various data sources within our database.

The database contains information on some 11.6 million individuals, all of whom have had at least one period in receipt of unemployed benefits since 6th April 2004 (the tax year ending 2005). 10.6 million have been matched to HMRC data, meaning that employment data from 1998/99 and earnings data from 2003/04 is at least partially available for them; 4.7 million have been matched to at least one learning aim in the ILR from 2002/03; and 5.7 million have been matched to at least one referral by Job Centre Plus staff in the LMS system from 2006/07 with some partial data for earlier years. Table 2 shows the volume of data recorded on the 11.6 million individuals for tax years[11] 2001 to 2013.

Table 2: Observed activity data, individuals receiving unemployed benefits since 6th April 2004

	Number of individuals ('000s)					
Tax year ending	in learning	in employment	receiving any benefits	receiving active benefits	with earnings data	with LMS referrals
2001		3639.9	2041.1	1250.1		
2002		4120.0	2320.6	1288.1		
2003	736.3	4452.9	2629.8	1372.9		
2004	1040.2	4769.7	2918.5	1444.6	2760.7	
2005	1128.5	5181.6	3188.7	1518.0	4862.7	
2006	1210.8	5567.1	4245.6	2127.7	5337.6	23.9
2007	1079.8	5801.5	4839.6	2294.6	5459.8	997.1
2008	1038.5	6067.2	5095.5	2180.3	5482.1	1112.2

[11] Tax years run from 6th April to 5th April.

2009	1133.6	6474.4	6189.9	2976.3	5696.0	1435.6
2010	1138.6	6294.6	6021.3	2886.3	5923.7	2050.2
2011	1117.1	6449.2	6386.1	3238.3	6320.2	2041.7
2012	1156.5	6555.7	7023.0	3725.7	6601.8	1934.0
2013	1196.3	6548.2	6429.1	3142.6	6480.3	1603.7

As with any research based on administrative data, an unknown level of error exists within the source data used. Firstly, some records may be missing. In the case of earnings and employment, data on the self-employed or those earning below the lower tax threshold may not be present. In the case of benefits and learning data, some records may have been managed manually and not recorded on administrative systems. In addition, some records may be in error (either wholly or partially) or be incorrectly linked. Both sources of error may affect groups of claimants differentially and therefore introduce an element of sample selection bias.

We would argue that the use of an achiever V non-achiever framework makes this less of a problem, as both our treatment and control groups select into treatment and, if anything, we would expect a higher proportion of these missing (likely lower) earners to be amongst the non-achievers. If anything their exclusion implies a downward bias in our estimates. A similar argument applies when we consider the fact that our data do not include all workers with income below the tax threshold (the fact that we do not observe returns for the self-employed is a limitation of most studies in this area).

The construction of Person-Keys within the database allows us to group overlapping and adjacent learning aims, and to do the same for periods of employment and benefit receipt. This makes possible an analysis of any of the three activities individually and collectively between any two dates or for any specified time period. From the ILR component, we identify the highest level of study within a learner's aims (if there was more than one) undertaken within a single learning spell (of which there could be more than one), together with its characteristics (such as length of the aim and whether the aim was completed or achieved. From the WPLS component, we create analogous spells for benefit periods, primarily to identify continuous periods (spans) when individuals are in receipt of active benefits (AB). Partially overlapping periods of Job Seekers Allowance (JSA), Training Allowance (TA) and Employment and Support Allowance Work-Related Activity Group (ESA-WRAG) could extend a continuous AB span but, unlike learning spells, breaks of more than one day in a continuous spell are treated as a separate AB span.

Some uncertainty exists within Benefit end dates: regular scans of operational databases are taken so the accuracy of these dates depends on the frequency of the scans. JSA scans, for example, are taken every 14 days. The employment spells we have created (from P45 data) should be accurate to the day since employers should in principle know when an employee begins and finishes a period of PAYE employment. However, 39% of spells appear to have start or end dates that are missing and have been assigned a default by HMRC.

The data and information available for the project are not sufficient to permit unambiguous procedures to replace missing or default data so that we can calculate precisely the length of employment spells. The volume of default data is such that we necessarily create an extensive decision rule process to, firstly, establish the possible range of start and end dates and, secondly, develop a structure to randomly assign dates within the ranges. Ranges are truncated by adjacent employment and out-of-work benefit spans[12]. Uncertain start and end dates for employment spells are then imputed within suitable 'gaps' in claimants' employment and out-of-work benefit histories.

Clearly there are likely problems that arise from this process of imputation when considering employment spell length, but again the question is whether this is likely to differentially impact treatment and control groups, in ways that we cannot accommodate. The omission of start or end dates in P45 returns is a result of employer behaviour, not the individual. It is quite possible that such employer behaviours are correlated with jobs characteristics (firm size and sector subject area for example), but it would seem reasonable to suggest that our controls for sector subject area, prior sustained employment[13] and a number of other characteristics should counter this. If we still believe that the distribution of imputed employment spells is not spread evenly across treatment and control groups, then our controlling for prior employment history should further accommodate this. Readers should refer to Bibby et. al. (2012) for more detail of the process of data matching.

2.2 Selection of the population for analysis

We first select *active benefits* spells where the claim start date falls within the inflow window, which covers April 2006 to 2008. These spells are then used to create a dataset containing a record for each individual[14], with the individual's claim start date marking the first reference point for analysis (or the claim start date of the first relevant unemployment spell, for those with multiple spells over the period). For each individual, the initial claim start date is considered as time (t) equal to zero, and then X (their expected date of referral to an ALMP) is calculated from this. We scan the Labour Market System (LMS), Individualised Learner Record (ILR) and New Deal (ND) datasets for all interventions/referrals (training or otherwise) that occur between time zero and X[15].

The overall population consists of 2.3 million individuals. We identify 1.52m individuals who have no referral to basic support and guidance (in the LMS) or evidence of fast-tracking to a ND (long-term unemployed) intervention. These individuals have not been flagged for support of any kind by Jobcentre Plus staff between claim start date and X. There are many possible reasons for this – they may have a very short spell of unemployment or have no obvious basic need in terms of skills/work-readiness[16]. Whatever the reason, we can clearly distinguish these individuals from the 0.73m who have been flagged for some form of basic intervention in the LMS, and this distinction is therefore an important component of our process of matching.

[12] An assumption is made that those in receipt of out-of-work benefits are not in employment concurrently. However, individuals may work up to 16 hours and still be entitled to JSA.

[13] We consider an employment spell to be "sustained" in cases where it lasts without interruptions for at least 6 months.

[14] Clearly some spells outside the window will need to be brought in for those with multiple spells, but with a relevant first claim start date towards the end of our inflow window. For instance, an individual with a first claim start date of Jan 28th 2008 and claim end date of March 28th 2008 would qualify for analysis, but a subsequent spell that started on April 15th 2008 would not qualify as a relevant spell [if selecting only on spells] but is a relevant spell as it is within the period between claim start date and X for this individual.

[15] In recognition of the potential margin for error around the expected claim start date of X, a 'fuzzy' X is created covering the period between X – 2 weeks and X + 2 weeks. Some scans of the data run to the start of this period (X – 2 weeks) and some run to X.

[16] It is also worth noting that the LMS may be incomplete, especially as we consider periods before 2006/2007.

The overall approach to evaluation, uses (i) ND and LMS data to differentiate distinct client groups, according to their apparent (flagged) need and (ii) ILR (FE) data, which provide more detailed information on achievement in training undertaken within FE settings, are used to differentiate treatment and control groups. Table 1 gives an indication of the learning aims and individual learners we observe in the ILR amongst the 2.3m unemployed individuals, between their claim start date and X (i.e. in the STU phase).

Table 1: Learning aims undertaken during the STU phase

	Duration (days)		%	Numbers	
Aim Type	Mean	St. Dev	Achieved	Aims	Learners
L1/L2- literacy	58	107	65%	195017	92823
L1/L2- numeracy	56	105	66%	185839	81261
L1/L2- ESOL	124	95	60%	41550	26506
Preparation for life and work	113	135	62%	73378	56111
Apprenticeships/ E2E	269	247	52%	43910	23275
Aims at level 2 or above of 480 GLH or more	410	269	40%	8478	7805
Aims at level 2 or above of 120 GLH or more	240	155	57%	58828	48488
ICT Aims at level 2 or above less than 120 GLH/ unknown GLH	91	111	59%	60252	28988
Other aims at level 2 or above less than 120 GLH/ unknown GLH	141	138	71%	134653	103566
Aims at level 1 or below	90	101	61%	85757	61103
Other aims (non-accredited, enrichment etc.)	58	84	81%	64170	44828

92,823 individuals have at least one ‘in-scope’ L1/L2 literacy learning aim and 81,261 have at least one ‘in-scope’ L1/L2 numeracy learning aim, amongst the entire population. In Section 3 we present initial results from a variety of analyses that provide estimate of the value added from L1/L2 Literacy and/or Numeracy achievement for the unemployed.

2.3 Econometric approach

In observational (non-experimental) studies the treatment group usually has different characteristics to those of the control group. To create robust estimates of any treatment effects, we require estimators capable of controlling for such differences. Standard regression-based approaches, that do not utilise data discontinuities or instruments, simply control for differences in characteristics by adding regressors. Matching methods account for any differences in characteristics between treated and control by matching each treated individual (achiever) to one or more controls (non-achievers); who are as similar as possible with respect to a given set of pre-treatment variables. Matching methods mainly rely on two crucial assumptions. First, the conditional independence assumption (CIA), which assumes that all the relevant differences between treatment and control are captured in their observable attributes. Second, the common support assumption, i.e. every achiever is assumed to have at least one counterpart in the control group. In recent years, a number of papers have highlighted the misapplication of matching methods by some researchers; thus, a new class of matching methods has emerged - dubbed “monotonic imbalance bounding (MIB)” (see Iacus et al., 2011) - that curtails the misuse of these techniques.

We implement one of these MIB methods, using coarsened exact matching (CEM). The idea of CEM is to temporarily 'coarsen' each conditioning variable into meaningful categories[17]; match exactly on these 'coarsened' data, and then retain only the original (uncoarsened) values of the matched data. 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[18]. Iacus et al. (2011) show that the CEM dominates commonly used existing matching methods in its ability to reduce imbalance, model dependence, estimation error, bias, variance, mean square error, and other criteria.

It is important to remember that the inherent trade-off of matching remains. With the CEM approach, larger bins (more coarsening) will result in fewer strata; fewer strata will result in more diverse observations within the same strata and, thus, higher imbalance (Blackwell et al., 2009). As recognised by Ho et al. (2007), matching methods are data-preprocessing techniques and analysts must still apply statistical estimators to the data after matching. Our estimates are produced using a CEM approach that:

- Matches (exactly[19]) on Number of months in Employment between Month (t) and Month(t-60); Number of months in Employment between Month (t-61) and Month (t-96); Number of months on Active Benefits between Month (t) and Month (t-60); Gender; and initial White/Non-White match for ethnicity; disability; prior caseworker referral; achievement/attendance of other lower level FE aims; local employment rate; age bands; months between claim start and course start[20].
- Then estimates a standard regression equation, using these matched (or re-weighted) achiever and non-achiever groups, controlling additionally for whether an unemployed individual has Children; a finer distinction of Ethnicity; whether individual is a Previous Offender; Age; ever Lone Parent; Number of prior LMS opportunities; Number of prior ILR (FE) aims started; more disaggregated labour market history variables.

Most of the variables we use as conditioning variables to justify the CIA - in particular, all variables that summarize individual employment histories - are calculated at the beginning of the unemployment spell rather than at program start. This reduces the potential impact of anticipation effects. In addition to detailed individual labour market histories, we have extensive information on personal characteristics, course attendance in FE institutions, and regional identifiers, which allows us to merge the unemployment rate and the index of multiple deprivation at the local authority

[17] For instance, if we are matching on previous earnings we may match on data that has been 'coarsened' by putting earnings into quartiles.
[18] Selecting matched samples reduces bias due to covariate differences, and regression analysis on those matched samples can adjust for small remaining differences and increase efficiency of our estimates (Stuart and Rubin, 2007).
[19] In addition, when making achiever V 'No FE' comparisons, we match exactly on whether individuals have a flag of need in the LMS.
[20] To maintain a sufficiently large number of matched cases for efficient estimations and coarsen our covariates into substantively meaningful categories, we proceeded in the following way: gender, disability dummy, white ethnicity dummy and prior caseworker referral are of course dichotomously split; then we split among unemployed achieving/dropping out/non-attending lower level courses, months between claim start and course start and age were split into tertiles, while local employment rate, active benefits and employment variables are split according to their median.

level. Our inclusion of dummies for the month of registration, helps capture seasonal unemployment effects.

Differently from Sianesi (2008) and Biewen et al. (2014) we do not have detailed profiles of job-seekers reported by the caseworkers. However, we do know if the caseworker considered the unemployed to be in need of some sort of basic support and guidance (as flagged in the LMS) and we use this information in the matching procedure. Together with the variables listed above, application of the CEM algorithm extensively reduces the imbalance between treatment and control group. Then we estimate a weighted regression equation, using these matched (or re-weighted) treatment and control groups, controlling additionally for the variables listed above.

The main concern is that there may be unobserved characteristics that simultaneously explain the particular treatment individuals received, and the treatment outcomes. However, a recent paper by Caliendo et al. (2014) confirms the importance of conditioning flexibly on lagged employment and wages, benefit receipt history, and local labour market conditions (also see Heckman & Smith, 1999). This study shows how rich administrative data including detailed labour market histories allow us to draw policy conclusions on the effectiveness of active labour market policies; as the addition of personality traits, job search and employment outlook, and socio-cultural characteristics does not substantially change the extent of the estimates when detailed labour market history variables are already included in the set of conditioning variables.

Despite the advantage of taking into account the dynamism of unemployed individuals' choice in taking a learning course, we do not adopt the dynamic treatment framework (see Sianesi, 2008; Fredriksson and Johansson, 2008) because of the difficulty of interpreting estimated effects[21] and of the requirement of additional distributional assumptions (Caliendo et al., 2014). Instead, we consider a classical static evaluation model and use the following definitions: participants are those unemployed who start a Literacy/Numeracy[22] aim within the STU phase of their unemployment spell; and nonparticipants are those who do not start a program in this period. Our achiever V non-achiever comparison is less likely to suffer from the problems that Sianesi attempts to overcome[23], but non-achievers may still drop-out as a result of securing employment.

[21] As a substantial fraction of nonparticipants in any given period might participate in a program shortly thereafter, the estimated effects are mixtures of the true program effect and differences due to shifted future program effects (Lechner et al., 2011).

[22] In cases where unemployed individuals attend Literacy/Numeracy courses more than once during the relevant unemployment spell, we evaluate the first course achieved and measure outcomes beginning with the first period after this first program start. In instances where the attender never achieves a Literacy/Numeracy course we take his/her first attendance as the course starting point.

[23] Nonparticipants so defined might be a positively selected subgroup of potential nonparticipants, since they are unlikely to enter a program because they have already found a job.

3. Results

The following discussion presents preliminary results from analysis of ILR-WPLS administrative data, focusing discussion on findings of the impact of literacy and numeracy interventions. All effects are measured monthly beginning with the month after the program started. This allows us to create graphs that clearly show the evolution of impacts over time.[24]

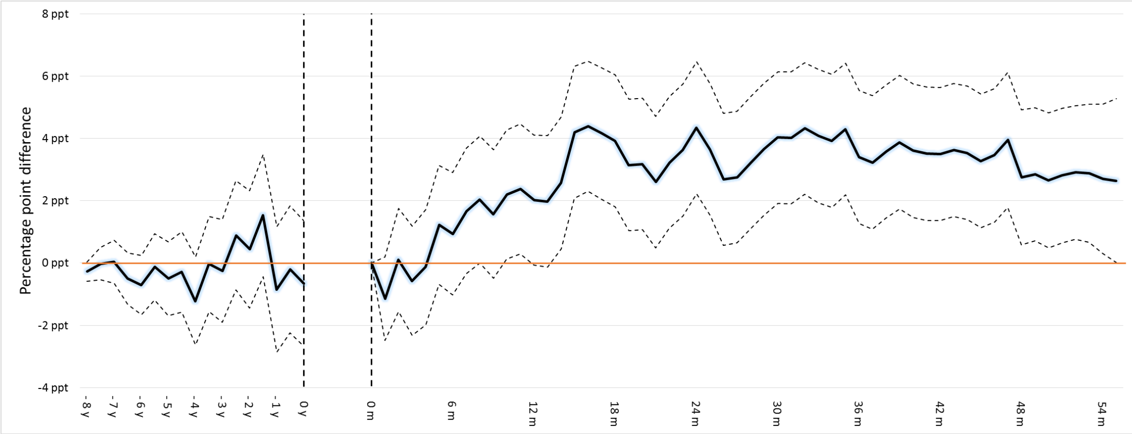
Chart 1 presents results from the first set of analysis for 18 to 24 year olds, comparing outcomes between (i) L1/L2 literacy and/or numeracy achievers with (ii) L1/L2 literacy and/or numeracy non-achievers/drop-outs. We consider the 'ATT' or Average Treatment on the Treated, as our estimate of value added, from comparison of achiever and non-achiever outcomes, as relevant to 'the kind of unemployed individual who we observe aiming for a L1/L2 literacy and/or numeracy qualification'. In the period between t-8 years and claim start date (t=0) we present graphically an idea of the quality of the match. If the matching procedure aligns treated and control as expected, there should be no statistically significant difference between the proportion of achievers in employment/sustained employment/active benefits and the proportion of matched dropouts, in any period prior to claim start. The thick black line, representing the outcome 'gap' between our matched treated and control group, remains close to zero for the entire period up to claim start date and well within our 99% confidence interval (the light dashed lines on either side).

In Chart 1 we observe an employment premium for 18 to 24 years old achievers over dropouts that, after a 1-year period where it almost monotonically increases, fluctuates around four percentage points. After 4 years this corresponds to an employment level of approximately 36%, against the 32% rate for non-achievers. An almost identical trend is reported in Panel B, meaning that the employment obtained is also more "stable" or sustained. Finally, Panel C shows how the probability of being on active benefits is higher in the first few months for achievers, but that this probability quickly turns negative and a lower proportion of achievers tend to stay on active benefits (however, the effect is rarely statistically significant).

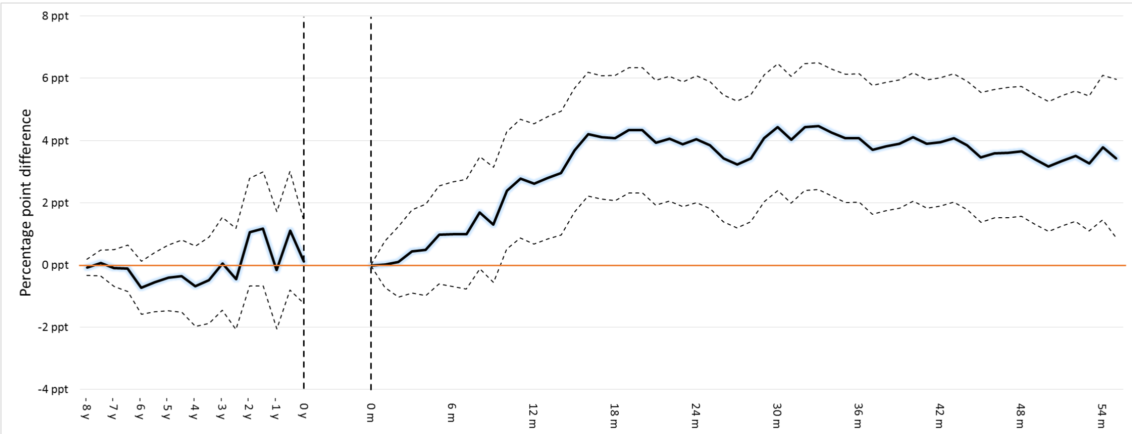
[24] Each graph displays on the horizontal axis the ATT, that is, the difference between the observed outcome with Literacy and/or Numeracy achievement and the estimated counterfactual outcome averaged over those who achieve the qualification in a given time-span. On the time axis, positive values denote months since course start, while negative values represent pre-unemployment months. We omit from the graph the period between the start of the unemployment spell and the start of the course in which both comparison and treatment individuals are unemployed.

Chart 1: Employment and active benefits outcomes for L1/L2 Literacy and/or Numeracy achievers, compared to non-achievers/dropouts: aged 18 to 24 amongst the STU

A: Employment



B: Sustained employment



C: Active benefits

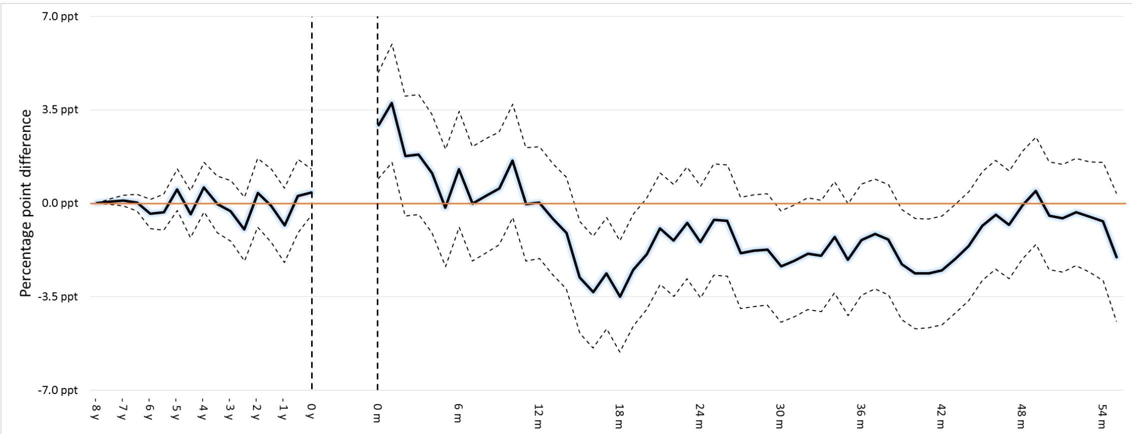
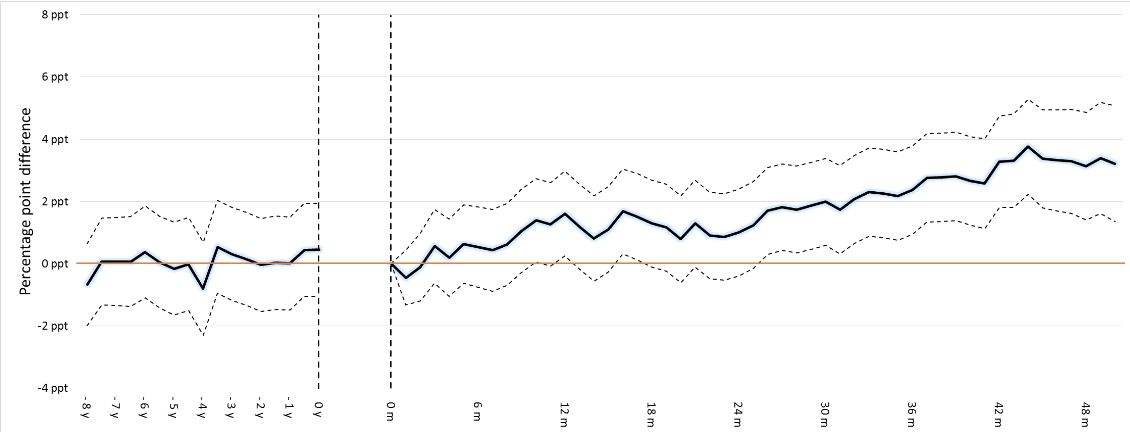


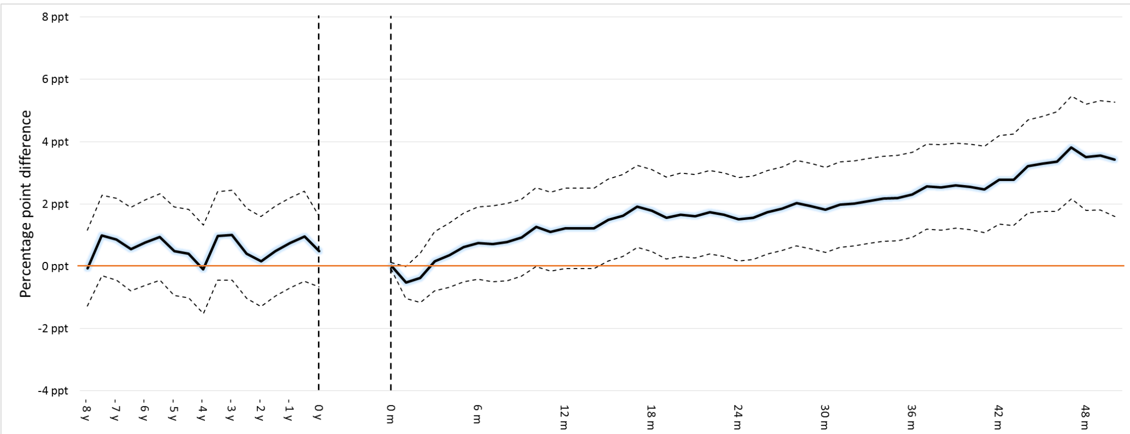
Chart 2 presents the estimated employment premium for all unemployed individuals aged 25+ who we observe achieving a L1/L2 literacy and/or numeracy aim, relative to those with the same aims, who do not achieve. Here we identify an employment premium for achievers over non-achievers, that becomes significant at points 12 and 18 months from learning start, and then between 25 and 30 months remains statistically significant at around 2 to 3 ppts until the end of our period of analysis. The employment premium for achievers steadily grows from 0 up to 3.5 percentage points after 4 years from course start date. Panel B shows a very similar pattern to that of Panel A and this once again suggests that the employment premium secured by those who achieve *L1/L2 Numeracy and/or Literacy* is not simply driven by differences in employment that is of a less substantial or temporary nature. If anything, we observe a slightly more substantial impact, as the difference in proportions of achievers and non-achievers becomes significant before 18 months (though it takes a little longer to reach 3 ppts). On the other hand, achievers aged 25+ are significantly more likely to stay on active benefits in the first year after course start (Panel C); though this effect quickly decreases towards zero.

Chart 2: Employment and active benefits outcomes for L1/L2 Literacy and/or Numeracy achievers, compared to non-achievers/dropouts: aged 25 to 55 amongst the STU

A: Employment



B: Sustained employment



C: Active benefits

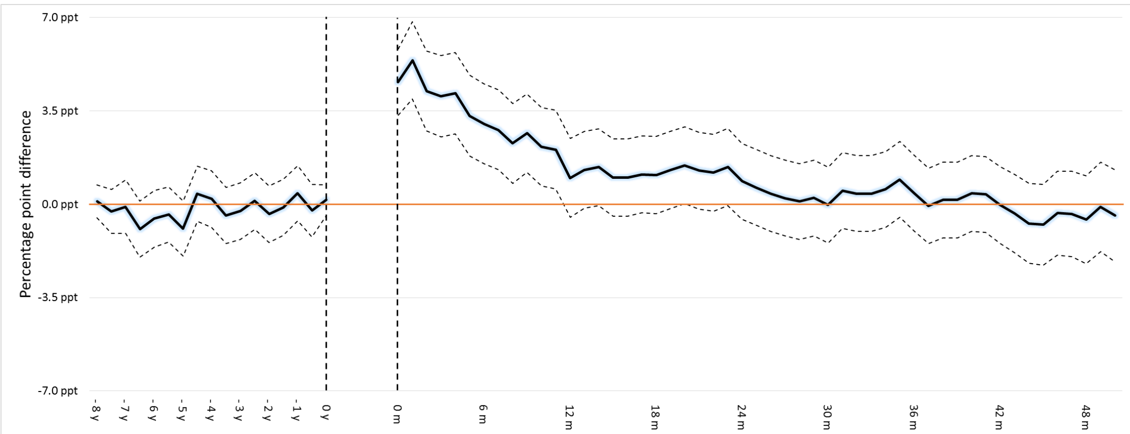


Table 1 presents summary measures that capture the outcomes from Charts 1 and 2. For instance, we observe a statistically significant 2 to 4 year average employment premium for achievers over non-achievers aged 18 to 24 of around 3.5 ppts; and around 2 ppts for those aged 25+. Overall this translates into a statistically significant employment premium of 2.4 ppts; as the 25+ age group undertake many more interventions than those aged 18 to 24, and therefore they 'weigh' more heavily in our weighted estimate (as one would expect, as the time between claim start date and X is much longer for the older age group)[25]. These premiums are slightly higher when we consider sustained employment outcomes, with the overall 2 to 4 year average estimate of impact rising to 2.7 ppts. However, we are only able to identify significant impacts (that average around -1.7 ppts) for those aged 18 to 24 when considering the gap between proportions of achievers and non-achievers on active benefits.

Table 1: Summary Labour Market Outcomes for *L1/L2 Literacy and/or Numeracy* Achievers, compared to Non-achievers: STU in the pre-2011 cohort

		Percentage Point Employment/Sustained Employment and Benefit Premium/gap in Years after Learning Spell Start				
		1st Year	**2nd Year**	**3rd Year**	**4th Year**	**2 to 4 year average**
Percentage Point Employment Premium	**Aged 18-24**:	0.010	0.034***	0.037***	0.035***	**0.035**
	S. E.	(0.007)	(0.008)	(0.008)	(0.008)	
	Aged 25+:	0.006	0.011**	0.019***	0.031***	**0.020**
	S.E.	(0.005)	(0.005)	(0.005)	(0.006)	
	All ages (weighted average)	0.007*	0.016***	0.024***	0.032***	**0.024**
		(0.004)	(0.004)	(0.004)	(0.005)	
Percentage Point Sustained Employment Premium	**Aged 18-24**:	0.012**	0.039***	0.040***	0.038***	**0.039**
	S. E.	(0.006)	(0.008)	(0.008)	(0.008)	
	Aged 25+:	0.006	0.016***	0.020***	0.030***	**0.022**
	S.E.	(0.004)	(0.005)	(0.005)	(0.006)	
	All ages (weighted average)	0.008**	0.023***	0.026***	0.032***	**0.027**
		(0.004)	(0.004)	(0.004)	(0.005)	

[25] We do not estimate one equation for all age groups as this is methodologically questionable (given that they have such very different expected dates of referral to ALMPs), but the overall weighted average of impacts from the two age groups is a measure that is appropriate for cost-benefit and policy analysis; and we provide an indication of its overall statistical significance.

		Percentage Point Employment/Sustained Employment and Benefit Premium/gap in Years after Learning Spell Start				
		1st Year	**2nd Year**	**3rd Year**	**4th Year**	**2 to 4 year average**
Percentage Point Active Benefit Gap	**Aged 18-24**:	0.010	-0.019**	-0.016**	-0.015**	**-0.017**
	S. E.	(0.008)	(0.008)	(0.008)	(0.007)	
	Aged 25+:	0.031***	0.012***	0.004	-0.002	0.003
	S.E.	(0.006)	(0.006)	(0.006)	(0.006)	
	All ages (weighted average)	0.025***	0.003	-0.002	-0.006	-0.000
		(0.005)	(0.005)	(0.005)	(0.005)	

Table 2 (which uses a 'No-FE' control group to estimate counterfactual outcomes) broadly confirms the findings on employment and sustained employment outcomes in Table 1 (which uses an achiever V non-achiever approach), in that we identify statistically significant premiums for achievers. For instance, we observe a statistically significant employment premium for achievers over the No- ILR (FE) group aged 18 to 24 of around 2.7 ppts; and around 4.8 ppts for those aged 25+. Overall this translates into a statistically significant employment premium of 3.9 ppts. These employment premiums are almost identical to those uncovered when we consider sustained employment outcomes in Table 2.

Table 2 suggests that those aged 25+ are securing much better employment and sustained employment returns, when compared to those aged 18 to 24. In contrast, when using an achiever V non-achiever approach in Table 1, it is the 18 to 24 year olds who secure a relatively higher return (for instance a 3.5 ppt employment return, compared to only 2.0 ppts for those aged 25+). This is something that we return to, as it provides some insight into the validity of achiever V non-achiever comparisons. However, for the present discussion, the main finding is of good employment and sustained employment returns to *L1/L2 Maths and/or English* learning in FE, whether we create counterfactual outcomes using a matched control group who do not undertake FE learning or those who select into FE learning, but do not achieve.

Unfortunately, when considering the active benefit gap between achievers and non-achievers for those aged 25+, the suggestion is that the former group are 3.8 ppts more likely to be on benefits between 2 and 4 years from the start of learning. This is in contrast to the weakly positive figure of 0.3 of a ppt in Table 1 for those aged 25+ and a [statistically significant] figure of -1.7 ppts for those aged 18 to 24 in the same table.

Table 2: Summary Labour Market Outcomes for *L1/L2 Literacy and/or Numeracy* Achievers, compared to *No ILR* groups: STU in the pre-2011 cohort

		Percentage Point Employment/Sustained Employment and Benefit Premium/gap in Years after Learning Spell Start ALTERNATIVE CONTROL				
		1st Year	**2nd Year**	**3rd Year**	**4th Year**	**2 to 4 year average**
Percentage Point Employment Premium	**Aged 18-24**:	0.015***	0.028***	0.028***	0.024***	**0.027**
	S. E.	(0.004)	(0.005)	(0.005)	(0.005)	
	Aged 25+:	0.050***	0.046***	0.046***	0.051***	**0.048**
	S.E.	(0.003)	(0.003)	(0.003)	(0.003)	
	All ages (weighted average)	0.036***	0.039***	0.039***	0.040***	**0.039**
		(0.003)	(0.003)	(0.003)	(0.003)	
Percentage Point Sustained Employment Premium	**Aged 18-24**:	0.011***	0.027***	0.026***	0.022***	**0.025**
	S. E.	(0.004)	(0.005)	(0.005)	(0.005)	
	Aged 25+:	0.041***	0.045***	0.044***	0.049***	**0.046**
	S.E.	(0.002)	(0.003)	(0.003)	(0.003)	
	All ages (weighted average)	0.029***	0.038***	0.037***	0.038***	**0.038**
		(0.002)	(0.002)	(0.002)	(0.003)	
Percentage Point Active Benefit Gap	**Aged 18-24**:	0.011**	0.004	0.004	0.005	0.000
	S. E.	(0.005)	(0.005)	(0.005)	(0.005)	
	Aged 25+:	0.069***	0.048***	0.035***	0.032***	**0.038**
	S.E.	(0.003)	(0.003)	(0.003)	(0.003)	
	All ages (weighted average)	0.046***	0.030***	0.023***	0.021***	**0.025**
		(0.003)	(0.003)	(0.003)	(0.003)	

Our analysis of the returns to FE learning amongst the unemployed using achiever V non-achiever comparisons uncovers good returns. Estimates of value added for 18 to 24 year olds, based on a comparison of outcomes between FE achievers and a matched control group who we do not see in FE learning, confirm these findings of good returns to FE learning. We see very little difference between counterfactual estimates created using an achiever v No ILR comparison, with those we get from an achiever V non-achiever comparison, when considering the 18-24 year age group.

Estimates of value added for those aged 25+, based on a comparison of outcomes between FE achievers and a matched control group who we do not see in FE learning, suggest even higher returns to FE learning**.** Estimates gained using an achiever V non-achiever comparison are lower than those when we use an achiever V No ILR (FE) comparison. For those aged 25+ we also see the estimates of benefit impact change substantially – with a suggestion that achievers are significantly more likely to be on benefits 2 to 4 years after learning, when compared to those who do not engage with FE during the period for analysis.

4. Conclusions

The outcomes presented here provide important new evidence on the returns to achievement to training in FE for the unemployed. They suggest that studies which fail to capture achievement of learning outcomes as part of any training undertaken by the unemployed will potentially under-estimate returns. The positive and significant employment and sustained employment findings are in contrast to many studies (Green et al., 1996; Caliendo et al., 2004) that find negative or insignificant returns.

We would suggest that the ability to differentiate achievers and non-achievers; the longer time period over which we can observe returns; the ability to differentiate populations of interest according to additional flags of basic need; the ability to match on up to 8 years of employment/unemployment histories; and the utilisation of more robust matching methods (because of the large numbers in our treated and control groups) goes some way to explain our uncovering of more favourable estimated returns.

The fact that we are able to compare estimates obtained using an achiever V non-achiever approach, with those obtained by comparing achievers with those who do not undertake FE learning, also allows us to refute one of the main previous challenges to the achiever V non-achiever approach to estimation using ILR-WPLS data (for instance in Patrignani and Conlon, 2011; Buscha and Urwin, 2013; Bibby et. al. 2014). Using an achiever V non-achiever approach it was possible (though highly unlikely given the evidence already amassed) that higher estimated impacts were a result of non-achievers experiencing one-off negative impacts that over-inflated estimated earnings/employment returns. This now seems highly unlikely, as our results for 18 to 24 year olds using a *No FE* control group, are very close to those secured using the achiever V non-achiever comparison; and for those aged 25+ they are actually higher (though this does not apply to benefit impacts).

Clearly when we consider the difference in findings from an (i) achiever V non-achiever and (ii) achiever V No ILR comparison for those aged 25+, a complicated picture emerges; the estimated returns actually rise when we compare to a No ILR group. This provides some insight into what was possibly happening in previous survey-based studies. Survey-based studies had suggested that some vocational qualifications (mainly taken within FE) at Level 2 were associated with negligible, or even negative, earnings returns[26]. Using ILR-WPLS data and an achiever V non-achiever/drop-out approach, Bibby et. al. (2014) find that FE qualifications are associated with good earnings and employment returns, and provide compelling evidence that the previous less favourable findings at Level 2 were a result of data limitations, rather than truly insignificant value added.

Individuals who hold a Level 2 vocational qualification as their highest form of learning are a unique group, with relatively limited labour market prospects, and it is therefore particularly hard to identify an appropriate control group to estimate valid counterfactual outcomes. One way of overcoming this is to compare the labour market outcomes of those who achieve vocational Level 2 as their highest qualification, with a group that have similar labour market opportunities (i.e. they select into the same vocational Level 2 qualification as their highest aim, but do not achieve and/or drop out[27]). The main potential weakness of this approach is that there may be unobservable characteristics or events driving selection into achievement, that are also correlated with subsequent labour market outcomes. Bibby et. al. (2014) use Coarsened Exact Matching with difference-in-differences, together with additional dissections of the data, to allay concerns over truly 'one-off' unobservable impacts on non-achievers, and the findings here provide further support for this approach.

When considering those aged 18 to 24 who have an unemployment spell, whom we match on employment history, and also on the extent to which we see flags of need in the LMS, we find that both (i) the non-achiever group and (ii) the No ILR group provide similar counterfactual outcomes; implying that we don't have such (potentially unobservable) heterogeneity driving selection between the two control groups. We have individuals at similar stages of their career (aged 18 to 24) and we are able to capture differences between treatment and control that might influence outcomes – as a result our matching using two different control groups leads to very similar counterfactual estimates.

In contrast, those who we see engaged in FE learning from the group of unemployed aged 25 to 55, are a selection of individuals from a much more heterogeneous group, who can be at very different stages of their careers, having very different reasons for being unemployed and therefore widely varying labour market opportunities. In addition, for the majority of these individuals we are considering a period of 18 months between claim start date and X, as opposed to only 6 months for the 18-24 age group. When we match on labour market histories and other variables, we remove a lot of this heterogeneity, but we still see some difference in counterfactual outcomes when estimated using a No ILR group, as compared to the non-achiever group.

It would seem that the strength of our approach is not just based on the comparison of achievers and non-achievers, as we obtain positive and significant findings when comparing achievers and those with No ILR. The strength is also in the ability of administrative data to control for much of the apparent negative selection into FE that survey-based studies are not able to accommodate. This is especially so amongst older age groups (here aged 25+) where even the selection of those with no qualifications in survey-based studies as a control group, leaves a group of individuals with varied

[26] See for instance, Dearden et al. (2004); Greenwood et al. (2007); Dickerson and Vignoles (2007); McIntosh and Garrett (2009).
[27] See Jepsen, Troske and Coomes, P. (2014) for a similar approach.

labour market prospects, that seem generally better than those who select into FE, in ways that cannot be observed.

In contrast, the ILR-WPLS-ND-LMS admin data allows us to capture and control for much of these problems, but even with the admin data, selecting unemployed individuals and using an analysis that matches on up to 8 years of labour market history, we can see that unobservable impacts for the 25+ age group can still alter our findings if we do not compare to a group who similarly select into FE. This seems to come through most in our estimates of the impacts from FE learning on unemployment benefit dependency.

This study represents a significant contribution to both the policy and academic literatures; it confirms that FE learning produces good labour market outcomes for some of the most disadvantaged groups in the English labour market; and sheds light on the reasons why previous studies may not have uncovered such findings.

References

Bibby, D., Knight, T., Speckesser, S. and Thomson, D. (2012), "Measuring Further Education Outcomes Using Matched Administrative Data: Constructing a Research Database", RM Data Solutions and Institute for Employment Studies.

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*.

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): pp. 837-897.

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

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*.

Caliendo, M., Hujer, R. and Thomsen, S. (2004), "The Employment Effects of Job Creation Schemes in Germany: A Microeconometric Evaluation", IZA Discussion Paper 1512.

Card, D., Kluve, J. and Weber, A. (2010), "Active labour market policy evaluations: A meta-analysis", *Economic Journal*, 120: pp. 452-477.

Green, F., Hoskins, M. and Montgomery, S. (1996), "The effects of company training, further education and the youth training scheme on the earnings of young employees", *Oxford Bulletin of Economics and Statistics*, 58 (3): pp. 469-488.

Ho, D.E., Imai, E., King, G. and Stuart, E.A. (2007), "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference", *Political Analysis*, 15 (3): pp. 199-236.

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

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): pp. 95-121.

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*, European Economic Association, vol. 9(4)' pp. 742-784.

Martins, P.S. and Pessoa e Costa, S. (2014), "Reemployment and Substitution Effects from Increased Activation: Evidence from Times of Crisis", IZA Discussion Paper No. 8600.

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

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.