## GrubbD-2020 comments-DEASP: Working paper: Evaluation of JobPath outcomes for Q1 2016 participants March 2019

[GrubbD-2020 comments-DEASP: Working paper: Evaluation of JobPath outcomes for Q1 2016 participants March 2019 1](#_Toc32574262)

[General 1](#_Toc32574263)

[p.17 The JobPath service and Figure 11 (map showing 90 Seetec and Turas Nua provider locations). 2](#_Toc32574264)

[p.19 Personal progression plan 2](#_Toc32574265)

[p.23 IV. Literature Review 3](#_Toc32574266)

[p24-25 Evaluating contracted public employment services 4](#_Toc32574267)

[p26-28 Evaluations in a dynamic treatment environment 4](#_Toc32574268)

[p29 V. Data and Description 5](#_Toc32574269)

[p.41 Persistent longer durations 7](#_Toc32574270)

[p44 VI Evaluation Approach 7](#_Toc32574271)

[p52 VII. Results: labour market outcomes 7](#_Toc32574272)

[p56 VIII. Discussion 8](#_Toc32574273)

### General

This evaluation report nicely summarises some background to Ireland’s “Labour Activation” policies, and how referrals to JobPath are made by quasi-random assignment and data are analysed to estimate the impact of JobPath.

It gives little information about the Jobpath services themselves: Jobpath office assessment interviews and follow-up interviews with clients; the acquisition of local vacancies and other placement resources; the staff profile; the ability to sanction for non-attendance or refusal of jobn offers; Intreo and Jobpath contractors’ contacts with each other (for sanction-related or other reasons) and overlapping contacts with clients; Jobpath management structures; the two JobPath contractors’ strategies and how they have evolved, etc. That could be a different and interesting piece of research with policy implications, e.g. if Jobpath local office performance seems to depend on Intreo local office behaviour sanctioning nonattendance which has been an issue in Australia.

Given the discussion of the various biases that arise in estimating impacts despite random assignment at the time of referral, and the authors’ awareness of these biases and their efforts to minimise them, it is plausible that the remaining biases do not greatly undermine the operational usefulness of the impact estimates. Ireland has hit on a neat way of using random assignment as an operational tool for individual jobseeker referrals and at the same time generating impact estimates differentiated by cluster, or for various other subgroups of the total.

This framework can easily be adapted to generate impact estimates also by Jobpath local office. Recontracting with providers based on their estimated impact is a far better management strategy than only contracting for “payment by results”. In a larger country with several “prime provider” organisations, it could be used to reallocate business in an underperforming local office area to a different prime provider (i.e. the principle applied in Australia, based on probably more accurate estimates of impact than Australia uses). If the contracted services supplement (rather than entirely replace) public services, in a larger country it would also be possible to introduce the contracted services progressively (e.g. in one wave introducing private services in half the regions of the country, and a second wave two years later): giving additional insights into the impact of the contracted services including the threat effects and local labour market displacement effects, as well as the costs involved for public services additional to the payments to contractors.

This DEASP working paper has not been carefully reviewed and edited for publication. The comments here pass over some typos and similar. The comments here do note certain possible errors in the table or figure titles (i.e. where the title and/or a footnote may be giving false or inconsistent information about the definition/coverage of the data presented).

#### p.17 The JobPath service and Figure 11 (map showing 90 Seetec and Turas Nua provider locations).

The current paper reports impact estimates by cluster, but the same methodology can, subject to sample sizes, generate impact estimates by provider and provider detailed location. The local office structure might imply aggregation of part-time and Outreach offices with their parent offices’ data (e.g. if that reflects management structure or if clients can opt for one or the other type of office). The JLD already allows estimation of local offices’ impact controlling for the cluster profile of the caseload (or the underlying raw client data). It could be feasible to add descriptors of local offices’ catchment areas (e.g. rural/urban, local age profile & industry structure, average commuting times of local people in salaried employment), i.e. if the first estimate of office impacts shows a significant relationship with a catchment area variables, allowance for that could be made. The office-level impact estimates could be fed back to Seetec and Turas Nua management and any major difference in the aggregate performance of Seetec vs. Turas Nua could be observed and used as input to contract renewal decisions (along the lines of Australia’s Job Network and later iterations of it)(although in Australia, central government reassigns detailed locations in which the current local office is underperforming to one of the other organisations - in Ireland that wouldn’t be possible consistent with the north-south split of offices).

##### p.19 Personal progression plan

This includes

* An agreed set of skills training, education, and development goals and actions,
* An agreed set of potential employment related experience interventions,

One might guess that potential referrals are to publicly-funded (SOLAS) skills training and education, “employment-related experience interventions” might (or might not) be referrals to a local Community Employment scheme, etc. This part of the Personal Progression Plans could be constrained by depending on 3rd party funding and organisation of the training, employment-related experience, etc.

p.19 Table 7. The LR 2-3 years total seems low compared with the LR > 3 years total: if these were stocks of unemployed by duration the data would imply only 17% of the stock with duration > 2 years leaves the LR each year (LR 3-4 years = 25625\*0.87; LR 4-5 years = 25625\*0.87^2; etc. adds up to about 125,000 LR > 3 years). Perhaps by 2018 people on the LR 2-3 years have quite often recently finished a 12 month JobPath phase and therefore are less likely to referred.

P20/Figure 12 2016 Time Lag between Referral and Start Date. After a first peak around day 9, the echoes around day 16 and 25 may reflect attendance at a 2nd or 3rd scheduled interview, after missing the first or second one. It would be helpful to have some description of the procedures and warnings and sanctions for persistent nonattendance (JobPath providers would need to notify Intreo at some point; it might be Intreo rather than JobPath that issues a warning about loss of welfare payments, etc.).

p.20ff/Figures 13, 14 The reader is not informed at what point in time the outcome status (i.e. “Cancelled after starting”, “Completed” or “Started (but not yet completed)”) is recorded. Referring to the central para of p.20:

* Outcomes must be recorded at least 12 months after the start date (2016 Q1) (if it were sooner, the number with status “Completed” after 12 months on the LR would be zero).
* Outcomes must be recorded even somewhat more than 12 months after the start date. If it is 13 months after the start date, the second group with status “Completed” i.e. those “who secure employment after working with the provider” and “have completed phase 2 of JobPath” will include only people who secured employment in their first month on JobPath.
* But outcomes must be recorded less than 24 months after the start date (2016 Q1): by month 24, all people who entered employment before month 12 will have completed phase 2 so no-one will have the status “completed phase 1 of JobPath and are in employment and still in phase 2”.

The number reported as “Started (but not yet completed)” actually depend entirely on when the outcomes are recorded, but the reader is never given that information! p.21 informs the reader that “11,609 people in 2016 started on JobPath, but had yet to complete it by year end”. Possibly this means that reported outcomes refer to status at year-end? but at year-end no-one who started during 2016 Q1 would have literally completed 12 months in phase 1. The description on p.20-21 of these statistics in Figures 13 and 14 should be extensively rewritten for clarity - also cross-checking that the (clear) information in one sentence does not contradict the (clear) information in another sentence.

### p.23 IV. Literature Review

The literature review nowhere mentions threat and anticipation effects. The reported data show that individuals with recent employment experience (even though they are long-term on the LR) often drop their claim (or risk having it cancelled) in the approximately 10-day period between referral to JobPath and their scheduled first interview (see Table 11, Figure 12). However at the same time, the existence of JobPath will tend to increase the employment hazard of individuals who pass 12 months on the LR (cf. the headings in Table 7), since from their 13th month on the LR they face an increased risk that they will have to participate in employment services and programmes at short notice (or find a job at short notice). Examples from the literature would usefully highlight the potential significance of this effect, which is additional to the type of impact studied and documented in the rest of this paper.

p.24 “Hamilton (2002) exploits a random assignment to 11 mandatory welfare-to-work programmes across the U.S, finding employment-focused programmes more effective than education and training.”

* Hamilton (2002) didn’t exactly “exploit a random assignment” since there were multiple random assignments in different localities, and multiple studies based on those, - “Hamilton (2002) synthesizes the overall findings from the five-year NEWWS evaluation” (OLeary Straits Wandner, “Job training policy in the United States”).
* It might be better to add, after “education and training”, “..but ‘mixed approaches’ that included both job search and educational opportunities seemed to work best”.

OLeary et al. quote directly from Hamilton (2002):“employment-focused programs generally had larger effects on employment and earnings than did education-focused programs”. But they also state “The most effective program emerging from the NEWWS was the Portland program, a hybrid employment- and education-focused model”. Fitzpatrick, Christie, Mark "Evaluation in Action: Interviews With Expert Evaluators” state “…the National Evaluation of Welfare to Work Strategies (NEWWS)… also found that programs that adopted 'mixed approaches' that included both job search and educational opportunities seemed to work best (Hamilton, 2002)." (these two citations are on pages accessible via Google books).

O’Connell et al. (2011) (appears twice on p.24 and once on p.27)

🡪 the bibliography lists this under K, i.e. Kelly, E., McGuinness, S., O’Connell, P. (2011) What Can Active Labour Market Policies Do? ESRi. [Online] Available from: https://www.esri.ie/system/files/media/file-uploads/2015-07/EC001.pdf [Accessed 14 March 2019].). Therefore the text should call it “Kelly et al. (2011)”.

#### p24-25 Evaluating contracted public employment services

This section cites an observation that “the evidence is mixed and it is difficult to draw conclusions as the form and models of contracting differ markedly between countries”. This is partly true but one could also say that only Australia has a nationwide longstanding system of provision by contracted providers, with placement outcomes (net of regression controls) used to reallocate business, creating a quasi-market in which only the best performing contractors survive longer-term and this has worked well for years. The other studies relate to contracting with some form of payment by results not impact estimates driving a quasi-market (most contracts were too small-scale and short-term for a quasi-market setup to be practical, although the UK Work Programme was relatively large) - it is not surprising but also not very informative, that their evidence is “mixed”.

p.25 the 3rd paragraph includes spurious references “(29)”, “(30)”.

#### p26-28 Evaluations in a dynamic treatment environment

p.26 A sentence here has the only use of the term “quasi-random”:

* “The quasi-random referral process in JobPath means that very different jobseekers will receive the JobPath service at the same time.”

Several later phrases state the random nature of the referral process but with a caveat:

* “there is, within categories of claim duration, a substantial element of randomness to whether people are assigned to JobPath (p.47)
* “Given the degree of random assignment involved in referral to JobPath, it is to be expected that..” (p.48)
* “Jobseekers do not self-select into JobPath, and the administrative selection process is based on a stratified random sample based on duration.“ (p.50).

In a few places the referral process is described as “random” with no caveat:

* the use of stratified random sampling means it is difficult to predict, with any degree of success, who will be referred (p.47)
* the Department of Employment Affairs and Social Protection (DEASP) selects jobseekers on a random basis for referral to JobPath (p.56)

The nature of the non-randomness in the sample that finally commenced, as compared with the control group is quite well documented, notably with the explanations:

* While the stratified random selection identifies the sample of long-term unemployed people to be referred, it is possible that, at the level of local Intreo centres, the sample referred may not match the stratification of duration bands. Not every Intreo centre will be able to refer exactly the required number of long-term jobseekers in exactly the proportions that would correspond to a stratified random selection. / Furthermore, the stratified sampling generates a sample of jobseekers who are referred but do not necessarily commence the JobPath service (see Section III). As this evaluation measures the impact of receiving the JobPath service, being referred is a necessary but not sufficient condition for inclusion in t’he population of interest. When examining those who started JobPath in Q1 2016 (a subset of those referred) and those who were eligible but not referred, the distribution of the duration of the ongoing claim at that time is markedly different between the two groups. (p.45)
* At the point of commencement (not referral), JobPath participants have longer durations than in the comparison group and, consequentially, lower mean earnings in previous years. / The reweighting based on inverse probability of treatment gives two groups that are, on examination of key baseline covariates, well balanced. (p.50).

p.27 the Sianesi (2004) estimation approach as described here, “the control group or basis of comparison is those people who are jobseekers up until a given point in time and have not taken part in the programme at least up to then”, does nothing to correct for selection effects arising from time-variation in jobseeker circumstances, e.g. when jobseekers who are pregnant, ill, anticipating long-term caring responsibilities soon, etc. for that reason do not participate in training and also do not enter employment.

p.27 O’Connell (2011) does not appear in the references (it might mean O’Connell et al which is actually in the references under K (Kelly)!

p.27-28 Kelly et al . (2015) and Cronin et al. (2017) are described with treatment group/control group language, but again they do not appear to correct for selection effects arising from time-variation in jobseeker circumstances, e.g. a jobseeker might see a business opportunity and therefore start on BWEA and also run a successful business, but that is not evidence of BWEA has a causal impact.

### p29 V. Data and Description

p.29 Harmon, Morrin and Murphy 2011 🡪 This is a citation has three author names but the other citations of articles with 3 authors use the style “Kelly et al. (2015)” – ideally the treatment of 3-author articles would be standardised (OECD switched to citations with up to 3 names).

p.29 The JLD data linking administrative data sets, informed by the 2011 overview, with individual data “re-arranged as a series of episodes” (with episodes defined by employment, unemployment status, participation in a programme), is interestingly different from a panel data structure. This is the first time I’ve seen a description of the episode data structure (where the data is stored as “points of transition from one status to another”)(though I’m not an expert). The approach facilitates a focus on transitions and duration in each episode, flexibly defined (e.g. one could look at duration of participation in any education vs. duration of participation in the current course of education). The episode data structure is still only in principle a different technique for storing the same information, with the advantage of being much more compact for certain types of data. The episode data could include variables such as change of address, youngest child starts school, start/end of intermittent health conditions, etc. They could be processed into panel data form i.e. snapshots of the individual’s characteristics at a series of fixed dates.

p.30 “what appear to be data entry errors are excluded by dropping observations where: \* earnings per week were greater than €352…” This statement needs explanation/clarification. Average earnings in 2016 were about 700€ pw ([www.cso.ie/en/statistics/earnings/earningsandlabourcosts/](http://www.cso.ie/en/statistics/earnings/earningsandlabourcosts/)) there is no obvious reason for earnings above half this level to be data entry errors. The data include the earnings of individuals in full-time work and off all benefits (data collected by the Revenue Commissioners). Possibly the text could mention that this observation-dropping procedure is applied only in the year that qualifies the person as statistically “long-term unemployed”? (in administrative data in some countries a person with high earnings in a recent week can still be long-term unemployed, perhaps not in Ireland?).

p.31 “..the mean earnings of the control group are higher than the treatment group and the mean duration in days of unemployment for the treatment group is slightly higher than the control group.”

Earnings in the year before the quasi-random assignment (2016 Q1) averaged 1411€ for the control group and 698€ for the treatment group, a huge difference. The treatment group initially i.e. including all individuals with an “intention to treat” (i.e. notified of their referral to JobPath) must have had (nearly) the same earnings in the year before assignment. The individuals who were referred but did not commence (i.e. did not attend the first client interview, or perhaps not attend a second or third, see p.20) must have had relatively high average earnings in the previous year on average. The JLD could be used to tabulate pre-referral and post-referral data for this group.

The procedure actually followed by the authors involves comparing the group that actually commenced Jobpath with a subset of the control group selected to be as similar as possible (described on p.45: “it is necessary to reweight the two groups - those who received the service and those who were eligible but did not - to ensure the measurement of outcomes at a later stage is a reasonable comparison”). However, the dropping of claims before commencement is probably motivated by factors that range from high employability (e.g. personal access to local job vacancies) facilitating rapid access to work (which is partially proxied by the JLD variable “high earnings in the previous year”) to low employability (e.g. strong aversion to work) which probably has no effective proxy in the JLD. It seems possible that the reweighting of the control group data lowers the employability profile of the control group considerably more than the dropping of claims before commencement biases the employability profile of the treatment group, which would leave the treatment/control comparisons biased.

Policymakers do want estimates for the impact of JobPath services on those who commence. The impact of referral to JobPath on those who did NOT commence may be seen as the “threat effect” of referrals. This “threat effect” might be not a policy objective, and JobPath providers may not have much influence over it, and they are not rewarded for maximising it.

However, JobPath impacts should (also) be reported also on an ITT (“intention to treat”) basis, given that there are no obvious technical barriers to doing this with the JLD data. In medical trials, ITT analysis can be difficult to implement reliably because outcome data are incomplete for the individuals who dropped out from the medical treatment: despite this problem expert advice and official guidelines call for ITT to be implemented - see [www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/):

A better application of the ITT approach is possible if complete outcome data are available for all randomized subjects. .. Anyone who follows these principles intelligently and has a vision to minimize bias should not worry further about “intention to treat”… in most cases, missing data could also be dealt by using the last observation carried forward (LOCF) method, whereby the last available measurement for each individual at the time point prior to withdrawal from the study is retained in the analysis. / US Food and Drug Administration (FDA) guidelines for “The Format and Content of the Clinical and Statistical Sections of Applications” state that as a general rule, even if the applicant's preferred analysis is based on a reduced subset of the patients with data, there should be an additional “intent-to-treat” analysis using all randomized patients. The FDA guideline further explains that the results of a clinical trial should be assessed not only for the subset of patients who completed the study, but also for the entire patient population randomized (the ITT analysis).

p.32 “As evident in Table 12 this trend continues with the treatment group having a higher mean of social welfare payments in 2017 and significantly lower mean earnings in 2017.” The table title is “Table 12: Social welfare payment history of the Control and Treatment groups” The titles of Tables 11 and 12 nowhere state what year the data refer to, only the main text indicates that they refer to different years (the editing/production values at this point are atrocious).

p.33 “It is, of course, useful to track over time the number of people with durations over 12 months and to compare absolute levels for different age categories. In certain circumstances, however, this approach of deterministic grouping can be a somewhat blunt analytical instrument. For example, those with 11 months’ duration may be quite similar to people with 13 months’ duration but a strict categorisation by duration places them in separate categories.”

This comment seems spurious i.e. it is not obvious that categorisation into clusters is “less sensitive to marginal differences in individual characteristics” than categorisation into short vs. long-term unemployment status. The sensitivity is harder to visualise because the reason an individual is put in one cluster rather than another is obscure anyway. I would prefer to leave readers with just the description of the clustering algorithm (it “calculates the optimal number of clusters, so that each cluster is, to the greatest extent possible, internally consistent (individuals in the same cluster are similar to each other) and distinct from other clusters (individuals in one cluster are different from those in other clusters).”

p.35 Figure 16: Younger Casual Claimants age distribution

This Figure (like several but not all of the later Figures) shows implausible spikes in the age distribution i.e. apparent bunching at multiples of 5 (20, 25, 30, etc.). This could be because for many (or all?) individuals, age data are initially reported in 5-year boxes (“20-24” etc.) and subsequently ratcheted up each year. This could usefully be explained in a footnote.

##### p.41 Persistent longer durations

This group has the second largest JobPath population (Shorter Durations 121,932; Persistent Longer Durations 97,946; the other groups each have about 15,000 to 30,000). “This cohort is the farthest from the labour market, with only 56% having had an episode of employment in the past five years.” Thus it represents a large group on welfare payments but nearer to “out of the labour force” than to “employment” status (see further comment below on the impact of JobPath for this group).

### p44 VI Evaluation Approach

p.44-45 “The design of JobPath has to address a difficult challenge in identifying a control group with a reliable counterfactual outcome, and therefore estimating the impact of JobPath, for two main methodological reasons:

* All jobseekers may be referred to a JobPath provider immediately or at a later point in time, which rules out a straightforward comparison between those referred to JobPath and those not referred….
* The measured effect of JobPath is contingent on the time when referral occurred and the time that has elapsed since referral. For this reason, subsequent analysis will attempt to measure who does better – jobseekers referred soon after becoming eligible or jobseekers referred to JobPath long after passing the eligibility threshold.”

Regarding the first of these points, reported impacts compare the treatment group with the control group which “includes also those who received JobPath after Quarter 1 of 2016” (p.52), i.e. with no adjustment for the fact that part of the control group, at a later date, itself receives the treatment.

Plausibly the various biases that arise despite random assignment at the time of referral, given that the authors are aware of these biases and seek to minimise them, do not much undermine the operational usefulness of the impact estimates.

### p52 VII. Results: labour market outcomes

p52. This section is numbered incorrectly as “V. Results: labour market outcomes”

p.54 Tables 30-32.

* Are the table titles correct? According to the table titles, impacts on total (annual) earnings are reported for 2017 but impacts on earnings per week and welfare payments are reported for 2018.
* Are the table footnotes correct? Table 31 footnote reads “Includes all individuals with or without earnings in 2017” but the title says the data are for earnings (per week) in 2018.

p.54 In Tables 30-32 “Persistent Longer Duration” is the only group for which the (estimated) impact of JobPath is a fall in social welfare payments that exceeds the increase in earnings. This could be a desired outcome, if it represents cutting off benefits for claimants capable of work but not motivated to work or for claimants in undeclared work. But it could be an undesired outcome, if it represents a loss of income for people who are unable to work. And it could call for a policy response e.g. a NEET type “outreach” service that identifies individuals who have dropped a benefit claim with apparently no other income, checks their circumstances and where relevant calls in social workers, helps with an application for disability benefit, etc.

p.55 “Table 32 outlines the decrease for each cluster in the [welfare] payments made to those who received the JobPath service compared to those who did not.”

🡪 in fact an increase in welfare payments is reported for the three groups “Younger Professionals”, “Shorter Duration” and “Older with strong employment history”! This seems surprising for the first two groups where annual earnings Table reports increases of over 2500€ and by more than 50%. Some commentary on the near-zero falls and increases in social welfare payments would be welcome. If the annual earnings impacts are really reported for 2017 whereas the welfare payment impacts are reported for 2018 (in line with the current titles of Tables 30-32) a possible mechanism is that higher earnings in 2017 increased the individuals’ benefit entitlements (i.e. entitlement to Unemployment Benefit which is contributory) in 2018.

### p56 VIII. Discussion

p.56. “It is reasonable to infer that the increased employment activity attributable to receiving the JobPath service prevents a drift out of the labour force to inactivity by long-term unemployed people at a time when increasing the size of the labour force through increased participation is a strategic priority.”

🡪 as noted above, the impact estimates seem to show rather the opposite for the “Persistent Longer Duration” group. At the same time, reducing welfare payments to a target population may increase the motivation to enter work. And the strategic objective should be to increase employment, not participation without employment (i.e. unemployment).