Research
Discussion
Papers

PSI Research Discussion Paper 17

Effective Job Search Practice in the UK's Mandatory Welfare-to-Work Programme for Youth

Michael White

ResearchDiscussionPapers

psi

Policy Studies Institute

Abstract

Administrative data from the UK's main welfare-to-work programme for unemployed and disadvantaged youth is analysed to identify differences in practice between local delivery areas, and to assess their effects on off-welfare outcomes. The findings reveal important similarities in the nature of effective 'work first' practices between this programme and some US programmes, despite large differences in the welfare context and systems.

1. Introduction

The USA and the UK are the two countries that have to the greatest extent embraced welfare-to-work programmes with an emphasis on job search, or 'work first'. There is substantial evaluation evidence that points towards the efficacy of work-focused programmes and job search services in both countries. A review focusing on early experimental programmes offering job search services for UI claimants in the USA (Meyer, 1995) indicated that these were generally effective. Related findings are presented as part of a more extensive review by Gueron and Pauly (1991). The evaluation of the California GAIN programme for AFDC recipients (Riccio et al., 1994), also based on an experimental (random assignment) design, revealed a striking contrast between results from the Riverside work-first version of GAIN and those of the five other counties in this programme, and has been widely influential. In the UK, positive impacts were found in an experimental evaluation of mandatory job search interviews for unemployed claimants (White and Lakey, 1992; Dolton and O'Neill, 1996). Similar impacts have been estimated non-experimentally (see the next section) for the New Deal for Young People (NDYP), a programme with a substantial work-first component. NDYP constitutes the empirical focus for the present paper.

There has also been a growing interest in getting inside the "black box" of programme evaluation to develop a more detailed understanding of what may be called micro-policy, micro-design or micro-implementation (after Berman, 1978). Mead (1997), for example, argues that evaluations should be followed by comparative studies of the performance of administrative units to elicit policy lessons. Some earlier studies in the USA, of a partly qualitative nature, were orientated in this way. For example, Chadwin, Mitchell and Nightingale (1981) based their assessment of WIN (the Work Incentives Programme) on a cross-State analysis and on more intensive study of service delivery units within five high-performing and five low-performing States. Among their wide-ranging conclusions, they asserted memorably that "programme management matters". Recently, interest in what makes programmes work has been connected to detailed analysis of area-based differences in programme governance, administration and delivery. This type of research seeks to identify processes or mechanisms by which programmes produce their effects, thus strengthening

confidence in the causal nature of those effects as well as providing more solid ideas for future programme design. An example is the paper by Riccio and Orenstein (1996) that developed further interpretation of the GAIN programme through analysis of richer, local data. A wider analysis of local programme characteristics, across several programmes, has also been produced by Bloom, Hill and Riccio (2001; 2003). Breaking into the black box and developing a more detailed analysis also raises complex issues of research methodology (see Mead, 1997; Heinrich and Lynn, 2000; Lynn, Heinrich and Hill, 2000). It is important to address these difficulties in order to make further progress in developing this field of inquiry.

The present paper seeks to contribute to these recent developments in three ways. First, it applies local-area analysis to data from the UK's largest welfare-to-work programme. This, the first such application in the UK, offers findings which can be set alongside those from the USA, thus increasing the generality of findings about programme delivery processes for welfare-to-work programmes[1]. Secondly, the present research is based entirely on the use of administrative data from the programme in question, rather than relying upon supplementary surveys to elicit information about local delivery practices. The paper illustrates how item analysis and scale construction procedures can be straightforwardly applied to administrative data defined at local level. Thirdly, the paper addresses some methodological issues which arise in examining micro-details of programmes at the local policy level, notably the presence of ambiguity, endogeneity, or measurement error in explanatory aggregate variables. As local-area and "black box" analysis is extended, it is important to incorporate methods of dealing with these issues.

The structure of the paper is as follows. The next section discusses related literature concerning programme variations. The third section describes the youth programme which is the focus of the research, and the data which was utilised. The fourth section discusses the methodological issues, and describes the research methods used. The fifth section presents the results, including sensitivity analyses. The findings are discussed, and conclusions are drawn, in the sixth and final section.

2. Concepts and previous research on local variations in programmes

A basic distinction is between 'macro-implementation' and 'micro-implementation' of programmes (Berman, 1978). Sandfort, Seefeldt, and Danziger (1998), applying this distinction, interpret macro-implementation as consisting of parameters established by high-level administrators "to structure the service delivery system". Micro-implementation, on the other hand, they interpret in terms of the decisions about service technology which are made at lower level within the front-line organizations that deliver the programme. Evidently, such lower-level decisions take place within a discretionary space which has either been purposely devolved to the front-line organizations, or which has been left vacant for them to colonize at their own initiative. Examples of macro-implementation variables within Sandfort and colleagues' study of PRWORA in Michigan, include the proportion of Work First providers by county and the proportion of non-profit agencies by county. Examples of their micro-implementation variables include the proportion of provider agencies imposing a requirement of immediate job search, and the proportion offering workshops to develop job search and job retention skills.

Both macro and micro variation can be further elaborated. All research in this field recognizes the need to take account of contextual variables reflecting economic and social conditions (a context important for macro-implementation decisions), and equally of the characteristics of clients (which are likely to influence micro-implementation). Recently, there has been growing interest in organizational issues around programmes, notably governance, management, and structural form (e.g., Heinrich, 2000; see also Considine and Lewis, 1999). These may involve both macro and micro elements, and are likely to be influential whatever the type of programme in view.

Any attempt to bring together all these levels is inherently complex. A near approach to such a synthesis is provided by Lynn, Heinrich and Hill (2000) who put forward a "reduced form logic of governance" for public management research. Their framework includes "structures" (incorporating a wide range of organizational

variables), "management", and "treatments" (all aspects of service definition and application), as well as environmental (contextual) and client characteristics. They argue that for service outcomes or outputs to be successfully modeled, all these types of variables need to be taken into account.   Even though their paper constitutes an unusually extensive review of relevant variables, the authors emphasize that they regard this framework as a preliminary step, and indeed it is possible to point to some simplifications in their account, vis-a-vis the previous literature.   Notably, their concept of treatments somewhat elides the distinction between macro- and micro-level programme design variables which has fruitfully been made by the authors cited earlier.

Research evidence on local variations in outcomes illustrates the foregoing concepts in terms of concrete variables. Illustrative of studies looking at governance-level variables are those of Jennings and Ewalt (1999), and Heinrich (2000). The former considered the coordination instruments used by Service Delivery Area (SDA) management, and whether or not JTPA organization was consolidated with that of Wagner-Peyser Act programmes. Both types of integration proved to have a positive bearing on several outcomes. Heinrich (2000) found a particularly clear influence of the use of performance incentives in contracting with service providers, as well as a number of variations between for-profit and non-profit services. Managerial variables have been explicitly considered in the follow-up investigations on GAIN and other programmes (Riccio and Orenstein, 1996; Bloom, Hill and Riccio, 2001; 2003). The former study, for instance, noted that job placement standards were used by Riverside's management as a criterion in performance assessment for caseworkers, while the latter examined the consistency of supervisors' and caseworkers' interpretation of the task (which presumably reflects communicative and control dimensions of management).

Turning to 'treatment' variables at macro-design level, the Charleston Placement Demonstration systematically varied the amount and intensity of job search assistance: impacts corresponded to the programmemed intensity level (Meyer, 1995). Somewhat similarly, Heinrich (1998) reported positive effects from intensive services targeted on an especially disadvantaged local community. Another programme design dimension, of particular significance for job-search or 'work first' types, is the

voluntary or mandatory nature of participation (e.g., Riccio and Hasenfeld, 1996; Mead, 1997). Relatedly, Schiller (1999) showed how changes in welfare rolls at State level depended partly on the nature of waivers from federal regulations adopted by the States, which shaped their versions of welfare programme reform along a 'soft-tough' dimension.

Treatment variables can also differ significantly at the micro-implementation level. The between-county differences in the GAIN programme can largely be classified under this heading. Both the Riverside version of the programme, with its emphasis on work-first, and the Alameda version with its strong encouragement for clients to enter education, were consistent with the overall GAIN framework, but that framework was flexible. The extent of permitted discretion in such cases means that between-site variations cannot necessarily be reduced to a matter of service intensity, as they can operate along different dimensions. Riccio and Orenstein (1996) discriminate multi-dimensionally between GAIN sites in terms of personalized attention, and enforcement including through use of sanctions. In a more extensive JOBS analysis, Bloom, Hill and Riccio (2001; 2003) use a "quick job entry" scale, personalized attention, and monitoring among their discriminators. In the Michigan PRWORA study by Sandfort, Seefeldt and Danziger (1998), an emphasis on focused job search assistance was found to have a negative relationship to aggregate employment outcomes, while the use of workshops to enhance job search skills was found to have a positive relationship.

Empirical research therefore gives good support to the argument of Lynn, Heinrich and Hill (2000) that a wide range of variables needs to be taken into account in the modeling of variations in public service outcomes. This is not to say that a standard set of variables can be applied. The particular selection of variables and the way they are used to frame hypotheses remain matters for judgement in the context of the specific programme. None the less, the broad concepts outlined in this section were found helpful in developing the research strategy to be described below.

3. The programme and data sources

New Deal for Young People (NDYP) is one of the largest and certainly the most costly welfare-to-work programme in the UK. It commenced in April 1998. Its aim is to help young people (aged 18-24) who have been unemployed for six months, or who fall into a number of disadvantaged categories (disabled, ex-prisoner, etc.), to find jobs and to increase their employability. In the UK, welfare benefits for unemployed individuals are not conditional on insurance, and only a small minority of claimants have unemployment insurance eligibility. Moreover, in common with some other European countries, unemployment benefit claims are not time-limited. Moving clients off welfare therefore has large and long-lasting implications for public expenditure.

A central feature of NDYP is the provision of a caseworker service, known as Gateway, with a 'work first' focus. The caseworker, known as a Personal Adviser, carries out an assessment of needs and provides support and monitoring for the individual to find a job, over a period intended to continue for up to four months from the time of entry (in practice, this often extends to six months). Those participants who do not find a job then pass into one of a number of further programme 'options': a waged job with subsidy paid to the employer, an unwaged work experience placement in the non-profit sector or in public sector environmental projects, or a full-time classroom-based vocational education course. These further placements can continue for up to six months (or up to 12 months for classroom education). If a waged job has not been obtained by the end of the placement period, the individual is required to return to a 'follow-through' period of supervised job search, closely similar to the initial Gateway period. This follow-through period continues for up to 13 weeks. All stages of NDYP are mandatory for the eligible group (if they remain on welfare).

The presence of the work experience and education options in addition to the 'work first' component creates scope for varying interpretation. Are the options intended as a deterrent which enforces focused job search, since those failing to find paid work know that their situation will default to option participation?[2] Or is the aim rather to

sift out those clients unable to compete in the job market, and then remedy their disadvantage through work experience and education? The initial governance design of the programme deliberately left flexibility to the local agencies to adopt their own interpretations (Department for Education and Employment, 1997), so as to adapt to local conditions and local stakeholder preferences. This was in contrast to the highly centralised specification and control of previous welfare-to-work programmes. This local flexibility, coupled with the breadth of the programme itself, makes NDYP a particularly valuable opportunity for assessing the effects of micro-implementation choices. NDYP was delivered through about 140 local agencies, known as Units of Delivery (UODs), with similarities to the SDAs (service delivery areas) in US programmes.

NDYP also offered a new research opportunity in terms of data. During 1997-98, a New Deal Evaluation Database (NDED) was established. This database collects information from a networked computer system on which front-line staff record service transactions with jobseeking claimants. Staff also record individual characteristics that are used in assessing eligibility or in performance monitoring. The unit record on the database is the individual, with facility for multiple spell records within person. Locational information facilitates aggregation to the UOD and higher administrative levels. The present study uses a database extract for the period January 1998 through August 2000. A further source of administrative information is a set of unit cost measures for each UOD, prepared centrally by the public employment service, for the period January-December 1998, inclusive.

Several evaluation studies of NDYP have been conducted and all have found positive impacts of moderate size from the overall programme - around 5-9 percentage points reduction in welfare claiming (see Riley and Young, 2001; Blundell et al., 2002; Wilkinson, 2002; White and Riley, 2002). However, Dorsett (2001) is the only previous study to focus on administrative data from the NDED, and his study was confined to estimating the relative impacts of the work experience and educational options. None of these evaluation studies has addressed the issue of local variations in micro-implementation.

4. Methodological issues, measures, and design of the analysis

The existence of an extensive administrative database, including information on processes applied by front-line staff, confronts the researcher with the questions 'What measures should be selected?' and 'How can the selection be justified?'. In answer to these questions, the following initial criteria are proposed:

(a) the measures should exemplify programme features that are salient and offer scope for local variation in practice; and
(b) the measures should be coherent and as far as possible distinct from other programme features.

These criteria are related to general principles of face validity and construct validity that are widely used in questionnaire scale construction or test construction. If they are satisfied, these criteria ensure that measures are neither arbitrary ('definition by labelling') nor trivial. Criterion (b) more specifically guards against the potential threat that the measures may proxy policies other than those which they purport to represent. For example, Gittleman (2001) investigated the influence of State waiver policies on exit rates from welfare, but concluded that "anomalies in the results suggest that [waivers] are actually serving as a proxy for other state-level changes". This situation, which might be referred to as 'unintended proxying', may arise either as a case of omitted variable bias, or as a misinterpretion of the selected variable. Unintended proxying, resulting from correlation of the given variables with unobserved variables can only be avoided (though perhaps never completely) by coverage of all the chief programme features through the selected variables. To avoid misinterpretation, it is worth devoting effort to grounding and/or confirming the interpretation of key variables by statistical methods, especially the construction of multi-item composite variables. In constructing multi-item measures, patterns of covariance between items are used to ground the interpretation of the measures, since interpretations are more constrained by item combinations than by single item measures. Multi-item measures also serve to increase reliability, relative to single-item measures. An additional possibility with programme evaluation is to examine the

patterns of effects over time, drawing on previous research findings to form hypotheses about the timing of effects. For example, 'work first' practices should show their effects quite rapidly during the periods when they are being intensively applied (see Riccio et al., 1994), while educational provisions would be expected to show their effects with considerable delay (Couch, 1992; Hotz, Imbens and Klerman, 2000).

A futher threat in the analysis of welfare policies is that of the possible endogeneity of policy choices. This problem has been recognized in macro-modelling of aggregate unemployment, whether cross-nationally (e.g., Bellman and Jackman, 1996) or regionally (e.g., Calmfors and Skedinger, 1995), but has received less emphasis in area-level research. The labour market or welfare policies at various administrative levels (and also the budgetary allocations from the centre to devolved levels) will often be influenced by current or recent experience of unemployment or welfare caseloads at each level. Then, because of the tendency for aggregate outcomes to be influenced by unobserved variables that are persistent over time, attempts to model outcomes in terms of the policy choices are likely to be statistically biased[3]. A common econometric response to this type of problem is the instrumental variables (IV) methodology, whereby the endogeneity of an explanatory variable is removed through use of another variable which is related to it, but unrelated to the outcome variable (Wooldridge, 2002). For instance, Calmfors and Skedinger (1995) used the party political composition at regional level as an instrument for welfare expenditures: the assumption is that political allegiance affects budget choices but not (at least directly) outcomes in the labour market. In the type of situation considered here, a simple heuristic of whether or not endogeneity is likely to be present is whether or not the potential measure of local policy is uncorrelated with (or not predictable from) pre-programme values of the outcome variable (here, unemployment). A significant relationship is suggestive of endogeneity, because unemployment is likely to be influenced by unobserved local circumstances which are persistent over time.

5. Measures of micro-implementation

As a basis for systematic development of measures, the chief features of NDYP in relation to job-search activation were identified through a review of official documentation and from qualitative studies and reviews. These features can be summarised as follows. (1) The amount of personal contact with front-line caseworkers, relative to previous provision, was greatly increased, with lengthy initial interviews and numerous follow-up interviews and telephone contacts becoming the norm (whereas previously caseworker interviews had been brief and widely spaced). (2) The scope for sanctioning clients for non-compliance was extended under NDYP with its widened mandatory provisions. (3) Additional resources were provided so that PAs could refer clients to external specialist services. These included skills assessments, careers guidance, disability services, counselling, mentoring, and behavioral therapy. (4) The programme stressed the importance of giving scope to client choice. To assist in choosing post-Gateway options, clients could be offered short courses known as 'tasters' which permitted them to make informed choice based on a trial.

From the administrative database, operational variables were selected to represent these features of NDYP. All were constructed from the administrative database over the first six months of the programme (April-September 1998) by aggregation to the local area (UOD) level. The emphasis was on identifying actions that are in the control of the front-line caseworkers, rather than on steps taken by the clients or dependent on other agents. To clarify this point, the caseworker controls whether or not to refer the client to a basic skills assessment, whereas the occurrence of that assessment may depend on the service provider accepting the referral or on the client showing up for the appointment. It is the former rather than the latter variable that is considered in this analysis.

The initial set of variables was then submitted to factor analysis and reliability analysis procedures to establish groups of variables which were distinct from one another and statistically reliable. Those factors with satisfactory reliability were then regressed on the average local (UOD) unemployment rate for 1997 (the year before

the introduction of NDYP), to provide an initial indication of potential endogeneity (see previous section). Two demographic variables, the youth ethnic minority proportion and the youth disability proportion, were also used as regressors.

Table 1 shows the variables screened in this way and the results of screening. It is of interest that the coarser measures of micro-implementation - numbers of job submissions and of submissions to options - did not pass these tests, either being correlated with pre-programme unemployment levels or forming part of weak factors. Also, referrals to external services had low inter-correlations with one another and with other variables, and so had poor scale reliability.

Table 1 Micro-implementation item assessment

Item description	Retained, or reason for dropping
Number of interview actions	retained, factor JSE
Days between 1st and 2nd interview	retained, factor JSE
Number of sanctioning referrals	retained, factor JSE
Number of education taster referrals	retained, factor IOC
Number of work experience taster referrals	retained, factor IOC
Number of environment-work taster referrals	retained, factor IOC
Number of job submissions	correlated with unemployment
Number of submissions to option places	correlated with unemployment
Number of basic skills referrals	weak factor
Number of disability referrals	weak factor
Number of unsubsidised job submissions	weak factor
Days from ND start to Gateway start	weak factor & correlated with demographics
Days from ND start to option entry	weak factor & correlated with demographics
Number of subsidised job submissions	did not factor
Number of mentor referrals	did not factor

Note: 'unemployment' and 'demographics' variables refer to the pre-programme year.

Two measures passed all tests and were labelled 'job-search efficacy' (JSE) and 'individual option choice' (IOC).   The JSE measure was derived as a principal components factor score loading on the average number of New Deal advisory interviews per client, the time-lag between the first and second interview, and the frequency of referrals to adjudication for benefit sanctions, all of which were inter-correlated. It thus combines elements from features (1) and (2) of NDYP: intensive caseworker contact, and sanctioning for non-compliance. This measure is interpretable as strength of local policy orientation towards the 'work first' dimension of the programme.   The IOC measure was a factor score loading highly on three measures of frequency of referral to different kinds of 'taster' courses for post-Gateway options. It represents feature (4) of NDYP and is interpretable as strength of local policy orientation towards encouraging individual choice, and facilitating entry to education and work experience components of the programme. These two measures were derived as orthogonal components in a factor analysis and are statistically independent of one another by construction. Both measures were unrelated to pre-NDYP local unemployment and to minority and disabled proportions.

Although the administrative database did not yield a satisfactory measure of the use of external services to support the job-search process, an alternative was found in another data source. A unit cost measure of Gateway services per NDYP entrant was available for 1998, at the UOD level. Importantly, this measure of expenditure *excluded* the costs of caseworker interviews and related administration, as well as the costs of welfare payments.   Thus the measure chiefly reflected external services contracted or purchased on behalf of clients. A high unit cost of Gateway can therefore be interpreted as indicating the local orientation towards using external services. The unit cost measure was not significantly correlated with the pre-programme unemployment rate, and had only slight overlap with the other two measures (correlation 0.20 with JSE, 0.18 with IOC). However, the possibility of endogeneity cannot be excluded since the time-period over which costs were established overlapped with the period in which outcomes were measured in this research. It also seems likely that the unit costs contained a non-ignorable amount of measurement error (in particular, because the costing system was new, there could be some unreliability in cost centre coding). It was decided to use this measure, labelled

'external service costs' (ESC), and to deal with the possible presence of endogeneity and/or measurement error by instrumental variables.

6. Control variables

Other variables included in the analysis, which are treated as controls, are of three main types: individual, contextual, and organisational. The individual variables were: age, number of previous unemployed welfare claims, the duration of the welfare claim at entry to NDYP, disabled (dummy), minority group (nine dummies), eligibility criterion (13 dummies), sex, marital status, the interaction of sex and marital status, usual occupational group (10 dummies), and the interaction of sex and occupational group. Labour market contextual variables were defined both at the individual level and the local aggregate level. At individual level, these were: the pre-programme unemployment rate for the travel-to-work area of the individual's residence, the chronological week of entry to NDYP, and the square of this week number (these last two variables capture time-related variations in the labour market, such as cyclical effects and labour queues). At aggregate (local area) level, the contextual variables were: region (nine dummies), rural/urban dummy, economically active population size, pre-programme unemployment rate, and home-resident student rate. Note that two unemployment rates are defined, one representing the labour market conditions facing the individual, the other representing the average conditions for the UOD (the two measures are based on different spatial units and had a correlation of 0.5 at individual level). Two organizational measures were available. The 'delivery model' classified the delivery areas by contractual type: public service, private sector, and mixed or partnership contracts. Additionally, a dummy identified certain of the delivery areas (N=12) which were selected to pilot the programme in advance of the national launch date.

7. Method of analysis

The analysis of programme outcomes uses a database extract covering all clients who entered in the period between October 5 1998 and 26 February 1999. The data extract contained 73,652 individuals, followed for a minimum period of two years. The data

covered all 141 areas. It was decided to exclude the 12 smallest areas, since these tended to have outlying values (both low and high) on a number of the variables which would be likely to reduce the reliability of the analysis. After removing these, and a small proportion of cases with defective data, the cases for analysis numbered 72,422 (98.3 per cent of the original extract).

The analysis was in two stages. The purpose of the first stage was to derive an aggregate (area) outcome measure that was adjusted for individual-level control variables. The second stage sought to model this outcome at area level through the measures of micro-implementation, together with the aggregate-level controls. Instrumental variables (IV) estimation was used at the second stage to remove potential bias in the variable for unit cost of external services. Ignoring IV, the models can be represented as follows:

$$Y_i = \alpha_{1t} + X_i\beta_{it} + D_j\gamma_{jt} + \varepsilon_{it} \quad (1)$$

where Y is the outcome variable and X the vector of control variables defined over individuals i, D is the set of area dummies, Greek letters indicate parameters to be estimated, ε is the disturbance, and subscript t (t = 1 ... 18) refers to monthly time periods for which the model is separately estimated (see also below); and

$$g_{jt} = \alpha_{2t} + P_j\delta_{jt} + W_j\eta_{jt} + \nu_{jt} \quad (2)$$

where g_{jt} are the estimated area differences in outcomes (net of individual controls) from the first stage of the analysis (these corresponding to the γ_{jt} parameters of equation (1)), P is the vector of micro-implementation policy variables over areas, W the vector of control variables at area level, and ν the area-level disturbance term (which includes estimation error for the g_{jt}). The disturbance terms in (1) and (2) are assumed to be independent; this is the chief simplifying assumption by comparison with fitting of a linear hierarchical model. (Within-area correlations in (1), that are provided for explicitly in linear hierarchical models, can also be dealt with through the use of robust estimators in OLS and IV models: see below.)

Model 2 can be written in more detail as follows:

$$g_{jt} = \alpha_{2t} + \delta_{1\,jt}JSE + \delta_{2jt}IOC + \delta_{3jt}ESC^* + W\eta_{jt} + \nu_{jt} \quad (2a)$$

where the first three upper-case labels are the elements of P as previously defined, and the asterisk following ESC indicates that it is instrumented. Two instruments for ESC were used: the per-area proportion of individuals from ethnic minorities; and a dummy variable, formed by dichotomising the unit cost variable at its median point. The rationale for the instruments is as follows. From previous research it is known that, in the UK, minority youth tends to have low uptake of available labour market services, thus a high minority proportion tends to reduce unit costs[4]. However, after controlling for human capital and other characteristics, minority youth does not appear to have a lower employment rate following unemployment. Dichotomisation at the median was shown by Wald (1940) to be a method of removing measurement error bias, and Durbin (1953) showed that it is an instrument. Appendix 1 contains details of instrument test statistics, indicating that the instruments are strong and giving no suggestion of instrument invalidity.

The individual-level outcome (Y_i) was whether the individual was 'off welfare' at a given time-point, coded 1, or 'on welfare' at that time, coded 0. 'Off welfare' means neither receiving unemployment-based benefits nor receiving benefits under NDYP. This outcome measure takes account of recurrent unemployment or recurrent return to NDYP. The aggregate, area-level outcome (g_j) was the area's adjusted difference in off-welfare rates relative to other areas at a given time point. The Y_{it} and g_{jt} outcomes were computed for every month from first entry into NDYP up to month 18, through separate analyses.

For the individual-level analysis, linear probability modelling was used (i.e., OLS with a binary dependent variable). The mean out-of-welfare rate was below 20 per cent in months 1 and 2, and the results for these months must therefore be treated cautiously (Cox, 1970). However, for periods 3 through 18 they were always above 20 per cent. The use of linear probability modelling, rather than nonlinear regression, facilitates subsequent interpretation. The second stage of analysis, at the aggregate level, was carried out by two-stage least squares (2SLS). At both the first and second stage of analysis, a robust variance estimator (the Huber-White estimator) was used.

At the first stage, this estimator takes proper account of the within-area correlations. At both stages, the estimator also takes account of heteroskedasticity.

The availability of outcome measures over an 18-month period strengthens interpretation. As described earlier, job-search support is provided in principle during the first four (in practice up to six) months of NDYP, and in a further three-month period for participants who remained jobless following an education or work experience option. Access to 'taster' courses and referrals to external specialist services were concentrated within the initial four to six month period. Thus, assuming these forms of client treatment are effective, each should have a somewhat distinctive pattern of outcomes over time associated with it. Areas scoring higher on the JSE measure should have enhanced outcomes in or shortly after the initial six-month period, and also at a later stage (around months 12-15) when unsuccessful clients are returning for follow-through support. In between, however, the advantage of the high-JSE areas will be reduced: so effects over time can be expected to be bimodal. Areas with a high score on the IOC measure are diverting clients from 'work first' towards education and subsidised work experience options, and this should result in reduced off-welfare outcomes during the initial six-month period. Increasingly positive effects on outcomes could, however, take place in the following year provided that more informed client choices lead to a greater efficacy of the options entered. Predictions in regard to the ESC measure are more ambiguous. If referral to external specialist services mainly serves to provide remedial treatment for clients, it seems unlikely that this would produce a short-term effect, and the gain from high ESC would be delayed somewhat similarly as from high IOC. But if high ESC frees caseworkers from supporting individuals with the most severe difficulties and permits them to concentrate their efforts on supporting job-ready clients, then outcomes could be enhanced during the initial six-month period.

8. Results

Results for the first stage of analysis are not shown, since its purpose was simply to produce the estimated area-level differences in outcomes, adjusted for individual-level characteristics. However the extent of the area-level differences is of some policy

interest. Averaged over the 18 months that were analysed, the standard deviation of the area coefficients (i.e., of the adjusted area-level off-welfare rates) was 4.2 percentage points. This however varied over time, being higher in months 5 to 8 and lower both before and after. At month 6, the standard deviation was 6 percentage points, while the difference between areas at the 10th and 90th percentile points was 14.8 percentage points. The between-area differences were therefore substantial even after controlling for a large number of individual attributes, most of which in fact were statistically significant in the first-stage analysis.

The main interest resides in the second stage of analysis, where the micro-implementation measures are considered as effects on the adjusted area-level outcomes. The results concerning the micro-implementation measures are summarised in Figures 1-3 (details for the control variables are not reported here, for reasons of space, but full results can be obtained on application to the author).

In Figure 1, the results are shown for the measure of 'job search efficacy' (JSE). Areas scoring high on JSE had above-average performance on the off-welfare measure over the whole period. The pattern showed the expected bimodality, with the effect rising and becoming significant at 5 months from entry, then staying on a plateau for several months, then dipping somewhat, and then rising to another peak at 15 months from entry, which is just after the entry peak to the follow-through period for those still jobless. Thus, the interpretation of this measure as concerned with efficacious job search processing by caseworkers is broadly supported by the pattern of effects over time, in the context of this programme's structure. However, the effect of JSE was more persistent than expected, with a post-Gateway dip that was both delayed and brief: JSE was non-significant only in months 1-4 and 12-13. The simplest explanation is that this reflects variable rates of passage through the programme, with many individuals having extended stays in the Gateway process and many others leaving option placements early, and thus returning prematurely to the follow-through search process (see Bryson, Knight and White, 1999). These variations would tend to spread the effect of JSE over a wider time-span.
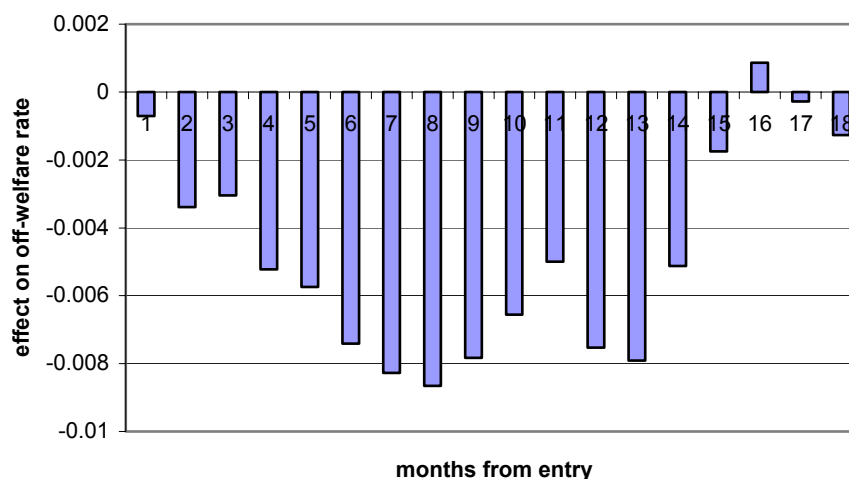
Figure 1 Effect of 'job search efficacy' (JSE) on off-welfare rate
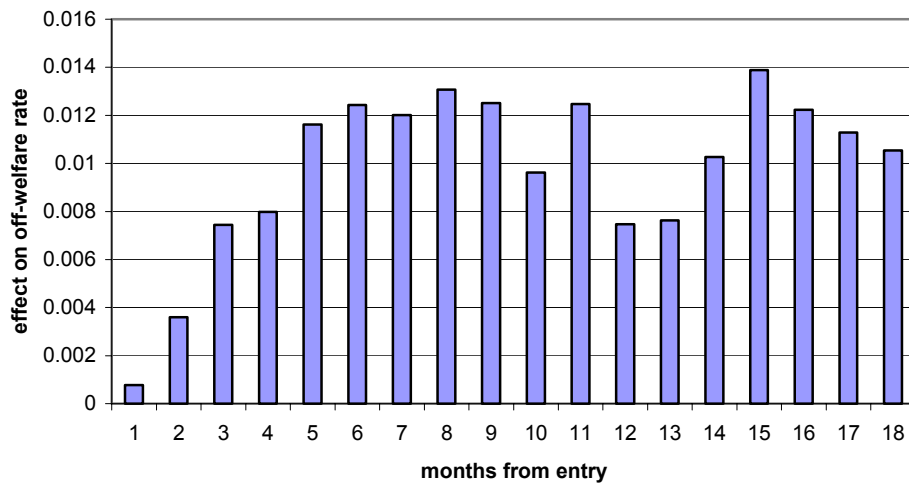
Notes: Coefficients from 2SLS analysis (each bar a separate analysis). The effect is per unit standard deviation of JSE. The coefficients are significantly different from zero (5 per cent significance level) in months 5-11 and 14-18.

Figure 2 summarizes the results for the measure of 'individual option choice' (IOC). Areas emphasising this aspect of the programme tended to have a depressed off-welfare rate in the initial period of intensive job-search, with relative outcomes becoming significantly negative in months 7-9. This is consistent with the interpretation that this approach tended to detract from the programme's 'work first' impact. However, the analysis showed no positive impact over the latter half of the observation period, so there was no evidence of a short-term payoff.
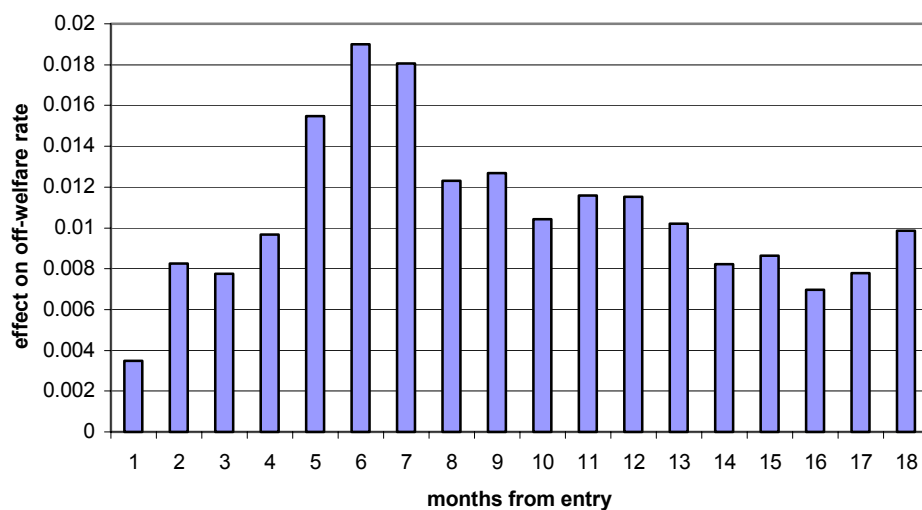
Figure 2 Effect of 'individual option choice' (IOC) on off-welfare rate

Note: Coefficients from 2SLS analysis (each bar a separate analysis). The effect is per unit standard deviation of IOC. The coefficients are significantly different from zero (5 per cent significance level) in months 7-9 and 12.

In Figure 3, the results are shown concerning the instrumented variable 'external service costs' (ESC*). Areas with higher expenditure on external services generally had somewhat above-average performance on the off-welfare measure, but this was significant only in months 5-7. Relative performance increased over the initial months to this peak level, and then fell back progressively although always remaining positive. In the context of this programme's structure, this pattern is most consistent with the interpretation that use of external services for clients with special difficulties freed time for front-line caseworkers to focus on supporting other clients' job-search. This is not to say that no remedial benefits were being obtained for the clients with difficulties. It is possible that these would not show through until a later time than that covered by this analysis.

Figure 3 Effect of 'external services cost' (ESC) on off-welfare rate

Coefficients from 2SLS analysis (each bar a separate analysis). The effect is per unit standard deviation of ESC (instrumented). The coefficients are significant (5 per cent significance level) in months 5-7.

To illustrate the magnitude of effects, coefficients and t-statistics for the micro-implementation variables are shown for months 6, 12 and 18 in the upper panel of Table 2. To facilitate comparison, the policy variables have been standardized and the effects are therefore per standard deviation (this also applies to Figures 1-3). Assuming normality, an area one s.d. above the mean on the JSE score had around a 2-3 percentage point advantage in off-welfare rate compared with an area one s.d below the mean, during the12 months when JSE was significant. An area one s.d. above the mean in expenditure on external services spent about £85 ($140) more per client than an area one s.d. below the mean. During the three months when the ESC variable was significant, the difference in off-welfare rates for these two areas was around 3.5 percentage points. For the IOC variable, the negative effects on outcomes were of smaller magnitude. In the four months when the IOC variable was significant, the outcome difference was about 1.5 percentage points between areas above and below the mean by one s.d.

9. Sensitivity analysis

Heinrich and Lynn (2000) have argued that the use of an aggregate-level analysis is particularly vulnerable to misspecification and thus to the production of misleading results. To avoid this, we have used a prior stage to adjust the aggregate-level dependent variable for compositional differences, have paid due attention to the issue of endogeneity in aggregate policy variables, and have used robust variance estimation. To assess robustness further, an alternative analysis was carried out, with individual data but with both the individual-level and aggregate-level regressors included in the specification (and without area dummies). Also, instead of instrumenting the ESC variable, it was included in its dichotomised form, ESCHI. According to Heinrich and Lynn (2000), OLS models at individual level with this range of regressors generally provide a reasonable approximation to a full hierarchical model. A set of analyses for outcomes at months 1-18 was obtained through linear probability models.

Overall, this alternative set of models produced results which closely matched the patterns described in the preceding section. Unsurprisingly the estimates for the ESCHI variable were about half as large as for the instrumented ESC variable, but significance was at a similar level. Overall, the results do not appear to be sensitive to the specification adopted. Illustrative numerical results from the alternative specification, for months 6, 12 and 18, are shown in the lower panel of Table 2.

Table 2 Illustrative estimates for the micro-implementation variables

(a)   Area-level aggregate analysis (2SLS): effects on off-welfare rate in percentage points (JSE, IOC and ESC* standardised)

	month 6		month 12		month 18	
	coeff.	|t|	coeff.	|t|	coeff.	|t|
Job search efficacy (JSE)	1.2	2.78	0.7	1.84	1.1	2.25
Individual option choice (IOC)	-0.7	1.80	-0.8	2.29	-0.1	0.43
External service cost (ESC*)	1.9	2.28	1.2	1.51	1.0	1.35

* Instrumented variable

(b) Individual-level analysis including both individual and area regressors (linear probability model): effects on off-welfare probability in percentage points (JSE and IOC standardised, ESCHI a dummy)

	month 6		month 12		month 18	
	coeff.	|t|	coeff.	|t|	coeff.	|t|
Job search efficacy (JSE)	1.3	2.71	0.8	1.89	0.9	2.30
Individual option choice (IOC)	-0.6	1.22	-0.5	1.25	-0.2	0.41
External service cost high (ESCHI)	2.2	2.38	1.2	1.73	1.0	1.74

10. Discussion

In this research, three variables representing variations in local micro-implementation of a UK programme for unemployed youth have been identified through

administrative data. Two were found to have significant positive effects on the local aggregate off-welfare rates while the third was found to have negative effects. All three variables exhibited patterns of effects over time which could be plausibly interpreted as relating to job search, given the structure of the programme. The findings pointed toward positive off-welfare outcomes resulting from a strong emphasis on 'work first' practices by front-line caseworkers. Practices contributing to positive outcomes were closely-spaced repeat interviewing of jobseekers, a large number of interviews (reflecting persistent follow-up), use of sanctions to enforce the mandatory nature of the programme, high expenditure on external services for clients (interpreted as a means of freeing caseworkers to focus on job-ready clients), and sparing usage of short courses that helped clients choose work experience or educational options.

These findings are broadly consistent with US evidence concerning the positive role of 'work first' micro-implementation practices. Indeed, the similarities to the GAIN evidence are particularly striking. This research shows that the efficacy of welfare-to-work practices is not peculiar to a national context. Despite the very different welfare systems of the USA and the UK, and the differently constituted client groups served, rather similar practices are found to be relevant to local reductions in welfare rolls. This is not to claim, of course, that practices of these types will be effective in every national context. The generality of the findings can only be extended through similar studies in other countries.

The combined effect of the three micro-implementation variables was greatest around the sixth or seventh month from entry to the programme, that is, at and just after the end of the initial period of assisted job search. Later on, as clients continuing their welfare spells entered work experience and educational options, the importance of the 'work first' practices naturally lessened, but they did not disappear even at 18 months from entry. This was very probably because an area's job search practices would also influence delivery of the subsequent 'follow through' component for those persistently unemployed.

The analysis did not find any evidence that local emphasis on encouraging client interest in work experience or classroom education, and on providing remedial

services for the hard-to-employ, improved outcomes in the latter half of the follow-up period. US experience, however, cautions against a hastily negative interpretion. These types of action may require periods of five or more years to show a return (Couch, 1992; Hotz, Imbens and Klerman, 2000), whereas the outcome data available for the present study extended for only 18 months. To assess developmental and remedial aspects of this (and other UK) programmes, one must await the availability of data over longer periods. It should also be stressed that this research did not directly consider the work experience and educational options. A study similar to the present one, but focusing on these options, would require information about the conduct of the options themselves, which is not at present available in the administrative database. Given such data, the methods used in the present study should be applicable.

Finally, this study has considered welfare outcomes rather than impacts, in common with much of the literature on local variations in programmes. This has the merit of connecting research with the kinds of measures most often used by programme administrators to monitor performance. None the less, there is a need to extend the approach adopted here to include local programme impacts in the future.

Appendix 1: Instrument tests

Instrumented variable: Unit cost of external services (ESC). Instruments excluded from second stage equation: dichotomised external services cost (ESCHI); standardised percentage of youth minorities in local population (ZETHPC).

F-statistic for excluded instruments (d.f. 2, 128): 43.26, P<0.001; partial R^2 for excluded instruments, 0.44.

Results for Hensen's overidentification test:

Month*	Chi-square (1 d.f.)	p
3	2.80	0.09

6	0.74	0.39
9	0.03	0.87
12	0.00	0.99
15	0.25	0.62
18	1.83	0.18

* The other 12 months (not shown) have chi-squares in the range 0.01-2.13, with probabilities in the range 0.14-0.96.

For discussion of tests, see Hayashi (2000), Wooldridge (2002).

References

Bellman, L. and Jackman, R. (1996) 'Aggregate Impact Analysis', in Schmid, G., O'Reilly, J. and Schömann, K. (eds.) *International Handbook of Labour Market Policy and Evaluation*, Cheltenham: Edward Elgar.

Berman, P. (1978) The Study of Macro- and Micro-Implementation, *Public Policy*, 26(2), 157-84.

Bloom, H.S., Hill, C.J. and Riccio, J. (2001) *Modeling the Performance of Welfare-to-Work Programmes: The Effects of Programme Management and Services, Economic Environment, and Client Characteristics*, Working Papers on Research Methodology, New York: Manpower Demonstration Research Corporation.

Bloom, H.S., Hill, C.J. and Riccio, J. (2003), Linking Programme Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments, *Journal of Policy Analysis and Management*, 22(4), 551-575.

Blundell, R., Costa Dias, M., Meghir, C. and Van Reenen, J. (2002) *Evaluating the Employment Impact of a Mandatory Job Search Programme*, Working Paper 01/20, London: The Institute for Fiscal Studies.

Bryson, A., Knight, G. and White, M. (2000) *New Deal for Young People: National Survey of Participants: Stage 1*, Employment Service Research Report ESR44, London: Department for Work and Pensions.

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*, London: Sage Publications.

Calmfors, L. and Skedinger, P. (1995) Does active labour-market policy increase employment? Theoretical considerations and some evidence from Sweden, *Oxford Review of Economic Policy*, vol. 11, 91-109.

Chadwin, M.L., Mitchell, J.J., and Nightingale, D.S. (1981) Reforming Welfare: Lesson from the WIN Experience, *Public Administration Review*, 41(3), 372-80.

Considine, M. and Lewis, J.M. (1999) Governance at Ground Level: The Frontline Bureaucrat in the Age of Markets and Networks, *Public Administration Review*, 59(6), 467-80.

Couch, K.A. (1992) New Evidence of the Long-Term Effects of Employment Training Programmes, *Journal of Labor Economics*, 10(4), 380-88.

Cox, D.R. (1970) *Analysis of Binary Data*, London: Methuen.

Department for Education and Employment. 1997, October. *Design of the New Deal for 18-24 year olds*. London: Department for Education and Employment / Welsh Office / The Scottish Office.

Dolton, P. J. and O'Neill, D. (1996) Unemployment duration and the Restart effect: some experimental evidence, *The Economic Journal* 106 (March), 387-400.

Dorsett, R. (2001) *The New Deal for Young People: relative effectiveness of the options in reducing male unemployment*, Discussion Paper, London: Policy Studies Institute.

Durbin, J. (1953) A Note on Regression when there is Extraneous Information about one of the Coefficients, *Journal of the American Statistical Association*, 48, 799-808.

Gittelman, M. 2001. Declining Caseloads: What Do the Dynamics of Welfare Participation Reveal? *Industrial Relations*, 40(4), 537-70.

Gueron, J.M. and Pauly, E. (1991) *From Welfare to Work*, New York: Russell Sage Foundation.

Hayashi, F. (2000) *Econometrics*, Princeton NJ: Princeton University Press.

Heinrich, C.J. (1998) Returns to Education and Training for the Highly Disadvantaged: What Does It Take to Make an Impact? *Evaluation Review*, 22(5), 637-67.

Heinrich, C.J. (2000) Organizational Form and Performance: An Empirical Investigation of Nonprofit and For-Profit Job-Training Service Providers, *Journal of Policy Analysis and Management*, 19(2), 233-61.

Heinrich, C.J. and Lynn, L.E. Jr. (2000) Means and Ends: A Comparative Study of Empirical Methods for Investigating Governance and Performance, *Journal of Public Administration Research and Theory*, 11(1), 109-38.

Hotz, V.J., Imbens, G.W. and Klerman, J.A. (2000) *The Long-term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Programme*, NBER Working Paper 8007, National Bureau of Economic Research.

Jennings, E.T. (1994) Building Bridges in the Intergovernmental Arena: Coordinating Employment and Training Programmes in the American States, *Public Administration Review*, 54(1), 52-60.

Jennings, E.T. and Ewalt, J-A.G. (1998) Interorganizational Coordination, Administrative Consolidation, and Policy Performance, *Public Administration Review*, 58(5), 417-28.

Lynn, L.E. Jr., Heinrich, C.J., and Hill, C.J. (2000) Studying Governance and Public Management: Challenges and Prospects *Journal of Public Administration Research and Theory*, 10(2), 233-61.

Mead, L.M. (1997) Optimizing JOBS: Evaluation versus Administration, *Public Administration Review*, 57(2), 113-23.

Meyer, B.D. (1995) Lessons from the US Unemployment Insurance Experiments, *Journal of Economic Literature*, 33, 91-131.

Riccio, J., Friedlander, D., and Freedman, S. with Farrell, M.E., Fellerath, V., Fox, S. and Lehman, D.J. (1994) *GAIN: Benefits, Costs and Three-Year Impacts of a Welfare-to-Work Programme*, New York: Manpower Demonstration Research Corporation.

Riccio, J. and Hasenfeld, Y. (1996) Enforcing Paticipation Mandate in a Welfare-to-Work Programme, *Social Service Review*, 70(4), 516-39.

Riccio, J. and Orenstein, A. (1996) Understanding Best Practices for Operating Welfare-to-Work Programmes, *Evaluation Review*, 20(1), 3-28.

Riley, R. and Young, G. (2001) *Does welfare-to-work policy increase employment? Evidence from the New Deal for Young People*, NIESR Discussion Paper 183, London: National Institute for Social and Economic Research.

Sandfort, J., Seefeldt, K. and Danziger, S. (1998) Exploring the Effect of Welfare Reform Implementation on the Attainment of Policy Goals: An Examination of Michigan's Counties, *Paper presented at the annual research conference of the Association for Public Policy Analysis and Management (APPAM),* New York City.

Wald, A. (1940) The fitting of straight lines if both variables are subject to error, *Annals of Mathematical Statistics*, 11, 284-300.

White, M. and Lakey, J. (1992) *The Restart Effect*, London: Policy Studies Institute and Employment Service.

White, M. and Riley, R. (2002) *Findings from the Macro evaluation of the New Deal for Young People*, Research Report No. 168, London: Department for Work and Pensions.

Wilkinson, D. (2002) *New Deal for Young People: Evaluation of Unemployment Flows*, PSI Discussion Paper, London: Policy Studies Institute.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data* (Cambridge, Ma.: MIT Press).

Endnotes

[1] Australia, New Zealand and the Netherlands have also moved towards a work-first approach in welfare-to-work programmes (Considine, 2001).

[2] The idea of deterrence was encapsulated in the phrase, often used by members of the UK government from the Prime Minister down, that in the programme there would be "no fifth option": meaning no inactivity (since NDYP offered four active options).

[3] Such bias is variously referred to as simultaneity bias or endogeneity bias.

[4] Consistent with this, in the present data the correlation of the youth minority proportion with the external service cost was -0.19.