NES Evidence Annex

The "what works" and comparison evidence base behind the National Employment Service (NES) proposal - the UK record on schemes that help people into work, and the closest schemes tried in other countries, each carrying an evidence grade. Supports the White Paper, the costing and the pilot design.

(The National Employment Service, or NES, is the scheme this institute is proposing: a guaranteed offer of paid, useful work for people who have been out of work for a long time. In this document, after it is spelled out the first time, it is referred to as "the NES".)

Evidence grades. Each finding below is graded for the strength of the evidence behind it. The grades run from A (strongest) to D (weakest):

- **A** - robust and causal: the finding comes from a randomised controlled trial (the gold-standard experiment, in which people are split at random into a group that gets the programme and a group that does not, so any difference between them must be caused by the programme) or a clean "natural experiment", and it has been repeated and confirmed.
- **B** - strong but not experimental: based on large amounts of real administrative records, or on a study design that mimics a proper experiment without quite being one.
- **C** - weak or indirect: a single study, a piece of modelling, or a contested result.
- **D** - contested or absent: little or no reliable evidence either way.

A note on two terms used throughout. "Quasi-experimental" means a study design that imitates a proper randomised experiment as closely as possible when a true experiment is not feasible. A "benefit-cost ratio" (sometimes shortened to BCR) is simply how much value you get back for every pound spent - so a ratio of £3 means £3 of value for each £1 of cost, and anything above £1 means the scheme returns more than it costs.

Part 1 - UK schemes that help people into work ("what works")

(These are all government programmes - sometimes called "active labour-market programmes", meaning schemes that actively help unemployed people find or prepare for work, rather than just paying them benefits. The table below summarises each one, what it achieved, whether it paid for itself, its evidence grade, and the lesson for the NES.)

A few abbreviations appear in the table and are explained here once:

- **The Department for Work and Pensions** (the government department responsible for benefits and employment support) is the source of much of this evidence; it is shortened to "DWP" in the sources list.
- "pp" means percentage points - so "+11pp" means the employment rate was 11 percentage points higher than it otherwise would have been.
- **The National Audit Office** (Parliament's spending watchdog, which checks whether public money is spent well) is shortened to "NAO".
- **Universal Credit** (the UK's main working-age benefit) is referred to in full.

- "Propensity-score matching" (referred to below by its initials, PSM, after being introduced here) is a statistical technique for comparing people who took part in a scheme with similar people who did not, in order to estimate the scheme's effect.
- "Difference-in-differences" (referred to below as DiD after this introduction) is a method that compares how an outcome changes over time for a group that got the scheme against a group that did not.
- "Payment-by-results" (referred to below as PbR) means paying the provider only when they achieve a result, such as getting someone into a job.
- "Jobseeker's Allowance" (an older unemployment benefit, referred to as JSA) and "Employment and Support Allowance" (a benefit for people whose health limits their ability to work, referred to as ESA) also appear.

Programme	Headline result	Did it pay for itself?	Grade	Lesson for the NES
The Future Jobs Fund (2009-11), 105,000 subsidised community-sector jobs	**+11 percentage points** more people in unsubsidised employment after 2 years; 7 percentage points fewer on benefits	Net benefit to society of **+ £7,750 per participant**; net cost to the Exchequer (the public purse) of **–£3,100**	**B** (using propensity-score matching, measured at 2 years)	Paid community-sector jobs *work*; the scheme was good for society overall but cost the public purse money. It was cancelled *before* its evaluation reported - the sequencing mistake to avoid.
The Work Programme (2011-17), about 2 million referrals, £2.9 billion	46 extra days in employment and 70 fewer days on benefit over 2 years	Returned £3.21 to society and £1.76 to the Exchequer for every £1 spent	**B/C** (propensity-score matching; only covers Jobseeker's Allowance claimants aged 25 and over)	**Payment-by-results led providers to "cream-skim"** - focusing on the easiest-to-help. The funding for the harder cases (people on Employment and Support Allowance) was about £550-690 each, against the £1,170 originally planned, far too low to serve them properly. Don't use payment-by-results for the placement phase.
The Work and Health Programme (2017 onwards), for people with disabilities or health conditions	The voluntary group: **+3 to 5 percentage points** more in employment; the mandatory (compulsory) group: **no significant effect**	About £1,560 per participant; no full benefit-cost ratio found	**A** (randomised controlled trial)	**Voluntary participation beats compulsion, even for a group with disabilities.** A dedicated personal adviser helps, but the size of the effect is modest.
Restart (2021 onwards), for the long-term unemployed	About 36% reached a job outcome (above its 31% target)	Social benefit-cost ratio revised down from **£3.80 to £2.44** once the number of participants halved	**C** (a group of participants compared with a similar comparison group)	Assuming high participant numbers, combined with fixed-cost contracts, is dangerous. Critics called it "glorified CV advice" - the NES's offer of actual paid work is what sets it apart.
Kickstart (2020-21), 163,000 youth placements, about £7,000 each	**+11 percentage points** in employment at 2 years; but **54%**	Returned £1.18 to society over 2 years, rising to **£3.15 over 5**	**B/C** (propensity-score matching plus difference-in-differences; the	Subsidising jobs in ordinary private companies carries **large deadweight** (paying for

Programme	Headline result	Did it pay for itself?	Grade	Lesson for the NES
	"deadweight" (more than half would have found work anyway)	**years**; to the Exchequer, £0.27 rising to **£0.49 over 5 years**	National Audit Office noted the "additionality" - how much extra was genuinely achieved - was *not measured*)	jobs that would have existed anyway); community and contribution roles have *structurally lower* deadweight because the work would not otherwise exist. Kickstart also couldn't reach the *economically inactive* - the very group the NES targets.
The Employment Retention and Advancement scheme (a demonstration project, 2003-07), a randomised controlled trial with 16,000 people randomly assigned	Lasting gains for long-term-unemployed men over **5 years**; gains faded for lone parents	For the long-term-unemployed-over-25 group, **£4.01 returned per £1**	**A** (randomised controlled trial)	**Support *after* a placement drives lasting gains** - and these only become visible over about 5 years. Treat the end of a placement as the start of a transition phase, not the exit.
The Sector-based Work Academy Programme (2011 onwards), a short scheme combining training and a work placement in a specific industry	**+13 percentage points** in employment at 2 years; biggest effect for people aged 50 and over	Returned £5.66 to society and £1.83 to the Exchequer per £1; cost about £428 per person	**B/C** (propensity-score matching plus difference-in-differences)	Cheap and effective **for those already close to being job-ready** - a useful fast-track for the NES's "bridge" group (people who need only a little help), but not a solution for those far from the labour market.

Part 2 - Comparison schemes in other countries

The two closest schemes to the NES are "job guarantees" - and both confirm the core idea works, with caveats:

(A "job guarantee" is a scheme that guarantees a job to anyone in a given group who wants one. "Living-wage" means paying enough to live on. "Non-market work" means useful work that the ordinary economy does not provide and that does not compete with existing businesses. "FTE" stands for full-time equivalent - one full-time job, or the equivalent made up of part-time roles.)

- **France - the "Territoires zéro chômeur de longue durée" scheme**, which translates as "Zero Long-Term Unemployment Zones". This is the closest comparison: voluntary, living-wage, socially-useful work outside the ordinary market, in economically depressed areas, paid for by redirecting the money that long-term unemployment would otherwise cost, with a strict rule against competing with existing businesses. Effect: **+74 percentage points** in employment for participants over 24 months (though this is largely circular - they are employed *by* the scheme). **Net cost of €11,000-€14,000 per full-time job per year** (the gross cost is about €28,000-€31,000); the scheme's own scientific committee found it **"far from budget neutral"** (in other words, it did not pay for itself). Grade **C+ to B–** (it uses a matched difference-in-differences design, with the areas chosen rather than randomly assigned, and small numbers; a fuller cost-benefit study was due in summer 2025). *Lessons:* the model works; **do not claim it pays for itself**; tailoring it to each local area and involving participants in designing the jobs both matter; moving people on into ordinary jobs was neither the aim nor something the study measured.
- **Austria - the Marienthal job guarantee** (also known as the MAGMA project). This is a genuine randomised controlled trial (registered in advance, with results forthcoming in the academic journal *American Economic Journal: Policy*): jobs guaranteed to everyone who wanted one, voluntary and well-paid, eliminated long-term unemployment in the district, raised income by **44%** (€1,280 a month versus €890), improved wellbeing across many measures, and did so **without displacing other jobs** - plus there was an **anticipation effect** (people's wellbeing rose simply from knowing the guarantee existed). Cost: **€29,841 per participant per year**. Grade **A** for internal validity (meaning we can trust that the scheme itself caused these results), but only **C** for external validity (meaning we cannot be sure the results would carry over elsewhere, because the study covered just 62 people in a single semi-rural village). *Lesson:* this is the strongest experimental evidence that the model works; it supports offering the guarantee to *everyone* within an eligible group; but whether it works at large scale is unproven.
- **Germany - the Hartz reforms. The key warning.** Public job-creation schemes (known in Germany as "ABM", short for *Arbeitsbeschaffungsmaßnahmen*, meaning subsidised work *outside* the regular labour market) produced **lasting harm to people's later chances of regular employment**, across decades of evidence (Grade **B**, based on extensive quasi-experimental study). The reason: sheltered, low-productivity placements let people's skills fade and carry a stigma compared with the open job market. *By contrast,* Germany's **dual vocational training system** (which combines classroom learning with paid work at a real employer) succeeds precisely because the work-based learning happens inside *real employers* with a structured path for advancement. *Lesson - the single most important design rule for the NES:* placements must build genuine skills, involve real employers, and offer a route into unsubsidised work; sheltered, dead-end work actively harms people who could otherwise be employed.
- **Denmark - "flexicurity"** (a system combining flexible hiring and firing with strong security and support for workers). Denmark spends about 2% of its national income on schemes that help people into work, and has high employment and short spells of unemployment - but this is the effect of a whole package, so the credit cannot be cleanly assigned to any one part (Grade **C**). *Lesson:* generous-but-conditional support and a real capacity to actively help people into work reinforce one another.
- **The European Union's Youth Guarantee.** Results are mixed and depend heavily on how it is carried out; it works best for the most disadvantaged; and it needs capacity built in advance so that the promised offer can actually be honoured within the time promised (Grade **C**). *Lesson:* a "guarantee" without funded, good-quality places to offer becomes an embarrassment.

- **United States subsidised and transitional jobs schemes** (including the National Supported Work demonstration, the Enhanced Transitional Jobs Demonstration, and the Subsidized and Transitional Employment Demonstration). Fifty years of randomised controlled trials (Grade **A**) show that transitional jobs reliably produce **modest** gains in employment and earnings in the short to medium term, largest for the most disadvantaged; traditional transitional jobs do better early on than schemes that simply subsidise an employer's wage bill; and the choice of who is targeted matters greatly. *Lesson:* the model reliably helps but does not transform people's prospects on its own - set realistic expectations.

Part 3 - Individual Placement and Support: the group furthest from work, and those with disabilities

(Individual Placement and Support, often shortened to IPS, is a specialist, well-tested method of helping people with significant health conditions into ordinary paid jobs. Rather than a long period of "getting ready" first, it places people in a real job quickly and then supports them in it.)

This has the strongest evidence base of any kind of employment support (around 30 randomised controlled trials and several reviews that pool many studies together) - Grade **A** for people with severe mental illness. Those who get Individual Placement and Support reach competitive (ordinary, open-market) employment at a rate of **55%, versus 25%** for those who do not - that is, they are about 1.63 times as likely to get a job. In routine UK practice the figure is about **41%** (with a median of 132 days to find a job). It is cost-effective: in the UK it returns **£1.41 for every £1 spent at 6 months, rising to £9.22 over 3 years**; and NHS savings of at least £20,000 per person over 5 years have been found in some groups. The cost of delivering it is **£1,157 to over £4,000 per participant per year** - and this depends on "fidelity", meaning how faithfully the method is followed (below about £4,000 per client, services fail to follow the method properly and the results collapse). The method works less well for opioid (drug) dependence and for people over 55. *Lesson:* the group furthest from work, and those with disabilities, should be offered an Individual Placement and Support element delivered at "Good" fidelity or better, budgeted at **£3,000-£5,000 per participant per year**, aiming for employment rates of 40-55% (for severe mental illness or disability) or 30-40% (for more common conditions) - with each person's needs assessed individually at the start, rather than treating everyone the same.

Part 4 - Valuing wellbeing (the largest "soft" figure in the cost-benefit case)

(Some benefits of a job are not financial - they are about how people feel. Economists try to put a money value on these so they can be weighed against costs. The standard unit is the "WELLBY", which stands for Wellbeing-Adjusted Life Year - one year of a meaningfully improved life. The figures below come from official UK Treasury guidance, known as the "Green Book", which sets out how government values costs and benefits.)

- **The non-financial wellbeing gain from being employed = +0.46 WELLBY per year** (from Clark and colleagues, *The Origins of Happiness*; based on tracking the same individuals over time, covering more than 100,000 people) - that is, the psychological benefit of having a job *over and above* the money it brings in. Grade **B** (this is a population average; individuals vary enormously - about half of those who are economically inactive show little loss of wellbeing).

- **The Green Book's value for one WELLBY = £13,000** (its 2019 central figure; the range is £10,000-£16,000; about £16,500 once updated for inflation). It is endorsed by the Treasury but how precise it is has been questioned - Grade **B/C**.
- → So the **evidence-based wellbeing benefit works out at about £5,980 per person per year** (0.46 × £13,000). This *replaces* the more optimistic £9,552 figure used in the first version of the cost model, and it is the single biggest reason the central estimate of society's benefit-cost ratio falls to roughly break-even (see Part 7).
- Unemployment shows **no "hedonic adaptation" and leaves "scarring"** - in plain terms, people do not get used to being unemployed (the harm keeps growing the longer it lasts) and the damage persists even after they find work again. This is a welfare argument for stepping in *before* unemployment becomes fully long-term. To avoid double-counting, use only the **non-financial** wellbeing figure here (the income side is counted separately in the financial figures).

Part 5 - Methods for evaluating the pilot (what to use, and what to avoid)

(This section is about how to measure whether the pilot actually works. Some of it is technical; the plain points are: design the evaluation before you start, compare against a fair "control" group, and measure effects rather than assuming them.)

- **Set up the evaluation before the first participant joins**, and **register in advance** the outcomes you will measure, how you will build the comparison group, and the statistical method ("estimator") you will use. (The Treasury's "Magenta Book", the official guide to evaluation, says a randomised controlled trial is the gold standard, and that a quasi-experimental design - one that mimics an experiment - is the "most defendable" alternative.) The failure of the Future Jobs Fund and Kickstart was *not measuring, in real time, how much extra they genuinely achieved* (as the National Audit Office found).
- **Do not rely on propensity-score matching as your main method** (King and Nielsen, 2019, showed it can actually *increase* the imbalance and bias it is meant to reduce). Instead use **entropy balancing** (a technique from Hainmueller, 2012, that is guaranteed to reduce bias and is "doubly robust", meaning it stays reliable even if part of the analysis is mis-specified) or "coarsened exact matching"; use propensity-score matching only as a cross-check.
- **Take advantage of rolling the scheme out region by region at different times.** Use the "Callaway-Sant'Anna (2021)" method for comparing groups over time (it avoids a known trap, in which the simpler "two-way fixed effects" approach can assign misleading negative weights to some comparisons); use a "synthetic control" (a constructed stand-in comparison area) where a whole area is rolled out at once; and use "regression discontinuity" (comparing people just above and just below a cut-off) wherever there is an eligibility-score or duration threshold.
- **Statistical power** (the ability to detect a real effect if there is one): to detect a roughly 5-percentage-point employment effect with 80% confidence, you need about **5,000-10,000 participants per group** (Kickstart detected an 11-percentage-point effect, give or take half a point, from 83,000 people). Make sure the design is large enough before you start.
- **Measure displacement and deadweight directly** - that is, jobs pushed out elsewhere, and jobs that would have existed anyway - using linked employer-and-employee records and control areas, rather than just assuming a figure.

- **Treat wellbeing as a main outcome to measure** - use the Office for National Statistics' four standard life-satisfaction questions (known as ONS4) at the start and at 6, 12 and 24 months, for both the treatment and control groups, so the result can be expressed in WELLBYs. (The Office for National Statistics, shortened to ONS, is the UK's official statistics agency.)

Part 6 - Cross-cutting lessons for the NES design

1. **Paid work beats advice** - programmes that give people an actual job (the Future Jobs Fund, Kickstart, the job guarantees) do better than advice-only support, especially for the economically inactive. The NES's core of real paid placements is the right instinct.
2. **The German warning shapes the split between permanent and transitional placements.** Sheltered work outside the ordinary market *harms* people who could be employed (as Germany's public job-creation schemes showed). So: offer **progression-focused placements that involve real employers and build skills** for those who *can* move towards ordinary work; and **reserve permanent placements for those who genuinely cannot** (people with severe disabilities; areas where the local job market has collapsed and there are no jobs) - exactly the "both, depending on circumstance" design, now backed by evidence, with the important rider that the default must be progression, not parking people indefinitely.
3. **Voluntary beats compulsory** (confirmed by the Work and Health Programme's randomised controlled trial).
4. **Don't use payment-by-results for the placement phase** (the Work Programme showed it diverts effort away from the hardest-to-help, who are exactly the point of the NES).
5. **Support after a placement drives lasting results** (the Employment Retention and Advancement scheme) - so fund a transition phase and expect a 5-year horizon before the full benefit shows up.
6. **Community and contribution roles carry structurally lower deadweight** than subsidising private-sector jobs (Kickstart's 54% deadweight is the warning) - a genuine advantage in adding real, additional activity, *as long as* the rule against competing with existing businesses is enforced.
7. **The job-guarantee model is proven in principle but unproven at large scale** (the French and Austrian schemes) - which is why the NES should be piloted first.
8. **Don't claim it pays for itself** (the French scheme's promoters did, and the claim did not survive scrutiny) - make the honest case instead: a real net cost, in exchange for real social value.

Part 7 - How this evidence shapes the cost-benefit case

Grounding the "soft" figures in the cost model in this evidence (wellbeing valued at 0.46 WELLBY × £13,000; the value of output anchored to the roughly £6,850 employer benefit found in the Future Jobs Fund) puts the **central estimate of the annual benefit-cost ratio to society at about 0.97** - essentially **break-even** (with a conservative estimate of 0.59 and an optimistic one of 1.46). Valuing wellbeing more generously would push it above 1, so the evidence-based figure is the conservative, honest one to report. The figures for the public purse are unchanged: the paying department spends about £15,800 per placement (about £4.75 billion at 300,000 placements), and the cost to government as a whole is about £13,800 per placement (about £4.15 billion). Two refinements the evidence forces on us:

- **The single-year, steady-state model understates the multi-year return for the *bridge* group** (those who need only a little help to move into ordinary work). The benefit-cost ratios for the Future Jobs Fund and Kickstart *rise* the longer you look (Kickstart's went from £1.18 to £3.15 over 5 years) because of the **tail of sustained employment after the placement ends**, which our single-year model does not credit (the Employment Retention and Advancement scheme shows this takes about 5 years to appear). So for those who go on to unsubsidised work, the true multi-year return to society is likely to be clearly positive.
- **For the *permanent-floor* group, the roughly break-even annual figure is the relevant one** - and the case for them rests on wellbeing, social value and dignity, not on a return from moving into ordinary work. This is where the value question (is it worth it?) is sharpest, and where it most clearly needs to be put to the public to decide.

Bottom line: the honest, evidence-based central case is roughly break-even for society within the year and a real cost to the public purse - turning clearly positive over a multi-year horizon *for the group that transitions into ordinary work*, and resting on social value *for the permanent group*. The way to resolve the remaining uncertainty is exactly the kind of pilot with a built-in evaluation described above, which measures the real effects on wellbeing, output, displacement and transition instead of assuming them.

Key figures (with grades)

(In the sources below, "IZA Discussion Paper" refers to a working paper published by the IZA Institute of Labor Economics, an independent German economics research institute.)

Parameter	Figure	Grade	Source
Future Jobs Fund: net benefit to society / cost to the public purse	+£7,750 / –£3,100 per participant	B	Department for Work and Pensions, 2012
Kickstart: benefit-cost ratio to society / to the public purse (5-year)	£3.15 / £0.49 per £1; 54% deadweight	B/C	Department for Work and Pensions Quantitative Impact Assessment, 2024
Work and Health Programme: voluntary employment effect	+3 to 5 percentage points (randomised controlled trial)	A	Department for Work and Pensions, 2023
Employment Retention and Advancement scheme (long-term-unemployed men): return	£4.01 per £1, sustained over 5 years	A	Department for Work and Pensions, Research Report 765
French "Zero Long-Term Unemployment Zones" scheme: net cost per full-time job per year	€11,000-€14,000 (gross about €28,000-€31,000)	C+	France Stratégie, 2024
Austrian Marienthal scheme: income gain / cost	+44% (€1,280 vs €890 a month) / €29,841 per person per year	A (internal validity) / C (external validity)	Kasy and Lehner, IZA Discussion Paper 16088
German public job-creation schemes: effect on later re-employment	persistently **negative**	B	IZA Discussion Paper 2100
Individual Placement and Support: competitive employment	55% vs 25% (about 1.63 times as likely); UK routine practice 41%	A	Campbell, 2019; British Journal of Psychiatry, 2024
Individual Placement and Support: UK return on investment	£1.41 (at 6 months) → £9.22 (at 3 years) per £1	B/C	Centre for Mental Health; Department for Work and Pensions Individual Placement and Support trial
Non-financial wellbeing gain from employment	+0.46 WELLBY per year	B	Clark and colleagues, The Origins of Happiness
Green Book value of one WELLBY	£13,000 (range £10,000-£16,000)	B/C	HM Treasury Green Book supplement, 2021
Propensity-score matching for building comparison groups	increases imbalance and bias - avoid	A (as a methodological finding)	King and Nielsen, 2019

Sources

(The "DWP" in some links below is the Department for Work and Pensions; "NAO" is the National Audit Office.)

UK: Future Jobs Fund (DWP 2012) · Work Programme evaluation 2020 · Work and Health Programme synthesis 2023 · Restart evaluation 2024 · NAO Restart 2022 · NAO Kickstart 2021 · Kickstart Quantitative Impact Assessment 2024 · Employment Retention and Advancement scheme, Research Report 765 · Sector-based Work Academy Programme Quantitative Impact Assessment 2025

International: OECD flexicurity Working Paper 139 · IZA Discussion Paper 2100, Hartz reforms / public job-creation schemes · France Stratégie / DARES, Zero Long-Term Unemployment Zones 2024 · Kasy and Lehner, Marienthal, IZA Discussion Paper 16088 · INET Oxford, Marienthal (MAGMA) policy briefing · International Labour Organization review of the EU Youth Guarantee · MDRC, National Supported Work demonstration · OPRE, Subsidized and Transitional Employment Demonstration, 8-year results · Georgetown, 50 years of subsidised employment

Individual Placement and Support / wellbeing / methods: Individual Placement and Support meta-analysis (Campbell, PMC6382127) · UK Individual Placement and Support routine practice (PMC12676244) · Centre for Mental Health, Individual Placement and Support · Green Book wellbeing supplement 2021 · King and Nielsen 2019 (propensity-score matching) · Hainmueller 2012, entropy balancing · Callaway and Sant'Anna 2021 · Magenta Book