NFER
National Foundation for Educational Research

Nuffield Foundation

Rethinking skills gaps and solutions

Technical Supplement Part B to Working Paper 4 of The Skills Imperative 2035

Dr Charalambos Kollias, Luke Bocock, Professor Bryony Hoskins, Dr Lisa Kuhn, Dr Frances Brill, Molly Jay, Dr Juan Manuel del Pozo Segura

Developing, piloting and validating a new instrument for the measurement of Essential Employment Skills

Dr Charalambos Kollias, Luke Bocock, Professor Bryony Hoskins, Dr Lisa Kuhn, Dr Frances Brill, Molly Jay, Dr Juan Manuel del Pozo Segura

NFER would also like to thank Verian (formerly Kantar Public) for their input into this Technical Supplement.

Published in June 2024
by the National Foundation for Educational Research,
The Mere, Upton Park, Slough, Berkshire SL1 2DQ
www.nfer.ac.uk

ISBN: 978-1-916567-10-8

How to cite this publication:
Kollias, C., Bocock, L., Kuhn, L., Hoskins, B., Brill, F., Jay, M., Segura, J. M. (2024). *Rethinking skills gaps and solutions. Technical Supplement Part B to Working Paper 4 of The Skills Imperative 2035: Essential skills for tomorrow's workforce.* Slough: NFER.

Contents

Nuffield Foundation

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

List of Figures

List of Tables

Glossary

Term	Definition
Essential Employment Skills (EES)	The six skills that are anticipated to be most heavily utilised across the labour market in 2035. These are transferable skills, specifically: communication, collaboration, problem-solving, organising, planning and prioritising work, creative thinking and information literacy.
Skills Supply	The level of EES that people – specifically workers, the long-term unemployed and young people – possess across the six skill domains, derived from self-assessments of their behaviours, on a scale of 0-100.
Skills Requirements	Refers to the EES people need to do their jobs, across the six skill domains, on a 0-100 scale, according to the results of our survey. They are calculated using people's self-assessments of the level and importance of each skill required to do their job.
Skills Gaps	Refers to the skills gaps calculated based on responses to our survey, from Skills Requirements minus Skills Supply for each skill domain.
Skills deficiencies	Where a worker (or group of workers) has a Skills Gap, and the Skills Requirements for their jobs are greater than their Skills Supply, according to workers' self-assessments.
Skills under-utilisation	Where a worker (or group of workers) has a Skills Gap, and their Skills Supply is greater than the Skills Requirements of their jobs, according to workers' self-assessments.
Standard Occupational Classification (SOC)	The SOC system is the main system for classifying occupational information in the UK. Jobs are classified by their skill level and context. The UK introduced this classification system in 1990 (SOC90). It has been revised every ten years, with the latest update taking place in 2020.
Occupational hierarchy	At its highest level of classification, the SOC (2020) classifies occupations into nine 'major' groups, based on skill level and skills specialisation. Occupations in SOC1 (Directors, managers and senior officials) typically require the highest skill levels, followed by SOC2 (Professional occupations) whereas occupations in SOC9 (Elementary occupations) typically require the least.
Higher skill-level occupations	These are occupations in the first three broad occupational groups (SOC1 to SOC3) in the SOC, specifically: 1. Directors, managers and senior officials (SOC1) 2. Professional occupations (SOC2) 3. Associate professional occupations (SOC3).

Mid- and lower-skill-level occupations	These are occupations in the bottom six broad occupational groups (SOC4 to SOC9) in the Standard Occupational Classification, specifically: 4. Administrative and secretarial occupations (SOC4) 5. Skilled trades occupations (SOC5) 6. Caring, leisure and other service occupations (SOC6) 7. Sales and customer service occupations (SOC7) 8. Process, plant and machine operatives (SOC8) 9. Elementary occupations (SOC9).

Purpose of this report

This report is designed to be read after *Rethinking skills gaps and solutions, Working Paper 4 of The Skills Imperative 2035: Essential skills for tomorrow's workforce.* Its purpose is to give further detail on the development, piloting and validation of our new instrument for the measurement of Essential Employment Skills. It is one of two technical supplements, the other one of which describes the analysis of data from our instrument and our results.

Introduction

In the last stage of *The Skills Imperative 2035: Essential skills for tomorrow's workforce,* we identified the set of skills anticipated to be most heavily utilised across the workforce in 2035 (Dickerson *et al.*, 2023). These were based on forecast changes in the relative importance of 161 different skill descriptors in O*NET[1]. Using these projections of the top skills in 2035 from O*NET, together with the findings of an earlier literature review (Taylor *et al.*, 2022), we identified six 'Essential Employment Skills' (EES): 1. Collaboration, 2. Communication, 3. Creative thinking, 4. Information literacy, 5. Organising, planning and prioritising, and 6. Problem solving and decision making. These EES are already a widespread constituent of existing employer-reported skills gaps. This suggests they may be a drag on economic growth, limiting individuals' employment and earnings opportunities as well as company's performance and productivity. However, we lack a detailed data-driven understanding of the current and anticipated supply of these skills, or the *Skills Gaps* that exist between workers' *Skills Supply* (i.e. the Essential Employment Skills workers possess) and their *Skills Requirements* (i.e. the skills they need to do their jobs effectively). **In this stage of The Skills Imperative 2035, we focus on measuring Skills Supply, Skills Requirements, and Skills Gaps in relation to Essential Employment Skills.**

Essential Employment Skills matter, but they are also inherently more inter-subjective and difficult to pin down and measure than their cognitive counterparts. Efforts to distinguish and measure these skills are complicated by the lack of consensus about which attributes actually constitute 'skills', with skills, personality traits, habits, attitudes, and commitment often bundled together and collectively referred to as "soft skills". Terms such as employability skills, essential skills, soft skills, life skills, citizenship skills and socio-emotional skills are defined inconsistently and used interchangeably, obscuring the nuanced differences in the aptitudes, attitudes, traits, values and behaviours to which they refer. This poses challenges for distinguishing and measuring Essential Employment Skills. **This Technical Supplement details how we responded to these challenges in the development, piloting, and validation of a novel new instrument to measure Skills Supply, Skills Requirements and Skills Gaps in relation to Essential Employment Skills.**

Our instrument – the NFER Essential Employment Skills survey – is the first of its kind. It measures people's Skills Supply in each of the six domains listed above, based on their self-reported behaviours and attitudes, as well as their 'Skills Requirements', based on the self-reported 'Level' and 'Importance' of each skill required to do their jobs. Skills Gaps are then calculated by subtracting Skills Supply from Skills Requirements. Our instrument also captures background information on respondents' demographic characteristics, industry, occupation, qualifications, employment status, socio-economic status, health status and training participation, as well as their salary, managerial status, and job and life satisfaction. This enables us to explore how Essential Employment Skills are distributed across the population, and how people's Skills Supply relates to their outcomes in the labour market and their satisfaction in life more generally.

This Technical Supplement details the development of our survey instrument, the survey design, data collection and weighting, instrument validation and how we projected future Skills Supply and Skills Gaps. A second Technical Supplement contains the results of analyses of our survey results.

This Technical Supplement comprises five sections:

[1] https://www.onetcenter.org/content.html

1. **Instrument development** – We first outline the development of our conceptual framework, the instrument specification, the development process, and initial qualitative trialling.
2. **Pilot and main stage survey design** – This describes the design of the pilots and main stage survey.
3. **Data collection and weighting** – In Section three, we set out the data collection approach, sampling, fieldwork, and survey weighting, for both the first pilot and the main stage survey.
4. **Instrument validation** - Section four details the validation of the instrument using Rasch measurement theory, covering both the first pilot and the main stage survey.
5. **Projecting future Skills Supply and Skills Gaps** – The final section describes the process we used to project future Skills Supply and Skills Requirements through to 2035.

1 Instrument development

1.1 Adopting a self-assessment approach to measuring people's 'Skills Supply' and 'Skills Requirements' in relation to Essential Employment Skills

1.1.1 Measuring respondents 'Skills Supply' of Essential Employment Skills using self-report methods

We define people's skills as the patterns of thoughts, feelings and behaviours individuals are able to exhibit in response to their environments, i.e. the behaviours they are able to exhibit when their situation demands it, which can differ from the behaviours they typically exhibit. There are no existing datasets that provide comprehensive information about people's level of EES. There are a number of potential methods for collecting information about people's behaviours. Our survey relies on self-report methods.

Self-assessment methods have been used to establish individuals' type and level of skills in large international studies such as the Survey of Adult Skills (PIAAC) (OECD, 2013). They are also frequently used in the field of personality psychology, one example being the 'Big Five' self-report personality questionnaire NEO-PIR (Costa and McCrae, 1992). Self-report measures can be a very useful way of establishing participants' perceptions and reflections of their own Skills Supply, especially where the evaluation of skills is inherently subjective (see Lucas, 2018). A range of existing self-report measures have previously been piloted and validated, which informed the items used in our scales, where appropriate. Self-assessment methods are also comparatively easy to administer, score and analyse, and altogether they provide a cost-effective way to compare individuals' Skills Supply across groups (Kline, 1993; Paulhus and Vazire, 2007).

Self-assessment methods enable us to gather data from the same sample of workers on both their Skills Supply and Skills Requirements, which enables us to equate the two onto a common latent scale and calculate Skills Gaps. We weight people's survey responses to correct for sample representativeness, and then re-weight our sample to account for anticipated changes in the composition of the population and the distribution of employment through to 2035. This enables us to examine how Skills Supply and Skills Gaps are likely to change in the next 10 to 15 years.

However, whilst the majority of published studies are based on self-reports, there are known shortcomings of these assessments. Self-reported attitudes or behaviours may not completely correspond with how participants would react, or feel, in reality. The measures may be prone to a range of biases, for example reference bias (Lira et al., 2022) which relates to differences in implicit standards held by individuals, or cultural biasing, threatening validity. Self-report measures are also vulnerable to social desirability bias as people may try to increase their 'moral worth' or show they are 'doing the right thing' (Sayer, 2007).

For self-report measures, it is also important to keep in mind that participants' responses often depend on context, such as question order, mode of delivery (phone or face-to-face assessment), current mood, day of the week or number of response categories (e.g. 7 or 11-point scale; ONS, 2018). Participants' responses in our survey may also depend on their employment context, particularly their industry and occupation, which could influence the norms they use as reference points when self-assessing their behaviours.

More innovative methods, such as situational judgement tasks, game-based assessments, and anchoring vignettes were considered for measuring respondents' EES. However, these alternatives are currently lacking a solid evidence base, for example on response consistency across participants (He *et al.*, 2017) and there is little research into how to produce valid vignettes (Grol-Prokopczyk, 2018), making those approaches less reliable.

More direct and explicit measures of participants' skills, such as the Animated Triangle Task which measures Theory of Mind, were also considered but the existing literature suggests they may have poor psychometric properties, such as poor discrimination power between participants.

The scales we developed to measure Skills Supply adopt a mix of frequency-based and agreement-based response scales to improve the functionality of self-report scales. This approach has been known to be better suited for measurement of the dynamic aspects of latent constructs (Tong, Bickmeier and Rogelberg, 2020) and improves scale performance (Marfeo *et al.*, 2014). Each scale contains between six and 12 items that ask about respondents' behaviours and attitudes, which are used to approximate their skill level. These statements will never fully capture the types and levels of skills people perform, both in-work and out-of-work and, consequently, a level of judgement is required when interpreting the results.

Our intention is not to suggest that workers' perspectives of Skills Supply and Skills Gaps should be given primacy over employers' perspectives – instead, our aim is to gather the missing worker perspective on skills gaps to add depth to our understanding of Skills Supply and Skills Gaps.

1.1.2 Measuring the 'Skills Requirements' of people's jobs using self-report methods

Whereas no existing instruments have been designed to measure the ESS people possess, survey items for measuring the Skills Requirements of people's jobs have already been used at scale by the Occupational Information Network (O*NET) in the US. O*NET profiles the tasks that are utilised in each of 964 occupations, and scores are provided for the 'level' and 'importance' of each skill utilised in that occupation, based on employers' and employees' assessments and expert assessor judgments. This metadata has been widely used by economists to understand how the demand for skills varies by occupation (e.g., Autor, Levy and Murnane, 2003; Deming, 2017). In the last stage of The Skills Imperative 2035, we mapped information from O*NET to the UK Standard Occupational Classification (SOC2020) and projected the future utilization of skills within each occupation using historic data from O*NET combined with our own employment projections (Dickerson *et al.*, 2023)

To measure Skills Requirements in the NFER Essential Employment Skills survey we use the 'level' and 'importance' self-assessment questions developed and used at scale by O*NET. Minimal changes were made to make the question wording and spelling suitable for an English audience. These ask respondents to rate the level and importance of each skill required in their current job. If they are not currently employed but have worked in the previous five years, we ask them to rate the level and importance of each skill required in their last job. By summing the importance and level scores for each skill for each respondent, we create a set of Skills Requirement scores for each respondent. We utilise the level scale anchors as O*NET to provide respondents with common reference points when rating the level of each skill required by their job. These anchors increase the objectivity of ratings. We equate respondents' Skills Supply and Skills Requirements scores onto a common latent scale using Rasch measurement theory. This enables us to make valid comparisons between Skills Supply and Skills Requirements (and hence also calculate Skills Gaps, based on the difference between the two).

1.1.3 Measuring personality traits

We also utilise our survey instrument to collect data on respondent's personality traits. This is not with the intention of analysing the relationship between individuals' EES and their personality traits in this stage of the research programme. Instead, our intention is to utilise this data later in The Skills Imperative 2035 research programme, when exploring the relationships between different types of cognitive and non-cognitive skills.

We conceptualise skills as the capabilities people can exhibit when the situation demands it, which are distinguishable from their typical behavioural tendencies (i.e. their personality traits). Steponavičius, Gress-Wright and Linzarini, 2023 explain:

> *An individual may have a habitual tendency to act in a particular manner, yet they might possess a high level of capability to behave differently when the situation demands it…someone may typically exhibit introverted and reserved traits, yet still possess the skill to assert themselves when necessary. Conversely, another person might display a talkative nature without demonstrating particularly skilled conversational abilities.*
>
> (Steponavičius, Gress-Wright and Linzarini, 2023).

To measure people's personality traits, we utilise the 'Big five' model, a widely acknowledged and empirically validated model for measuring personality traits and boiling them down into five core factors: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. For practical reasons, we use the "The Big Five Inventory–2 Extra-Short Form (BFI-2-XS)" version. This asks people to indicate the extent to which they agree or disagree with five sets of three statements (one set for each personality trait). People's responses are used to rank them on a scale for each personality trait, for example from 'closed' to 'open' on the openness trait. The scoring key is based on Soto and John, 2017. We equate people's scores for each personality trait with their Skills Supply and Skills Requirements scores, enabling us to compare all three.

We do not discuss the development of the conceptual frameworks and measures of Skills Requirements and personality traits further in this Technical Supplement because our items are drawn from well-established and widely used instruments. The focus of the remainder of this section is on the development of a conceptual framework and survey scales for the measurement of people's EES Skills Supply.

1.2 Developing a conceptual framework to underpin the scales for measuring people's Supply of Essential Employment Skills

This section sets out how NFER developed a conceptual framework to underpin the construction of survey items to measure the Supply of each of the six Essential Employment Skills. We provide an account of the process we followed to develop the conceptual development, define and describe each skill, and break each skill down into thematic areas. We start with a basic outline of the classification systems that were drawn upon to create the framework.

1.2.1 Review of existing frameworks and classifications

We started by examining the most relevant skills classifications and frameworks in the public domain. These included large-scale skills classification systems developed on national or supernational levels that are dedicated to capturing the skills of the entire labour market within their jurisdiction. Such classification systems offer categorisations of the skills needed for the majority of different job types. They may be used, for example, by policymakers investigating skills mismatches, by academics analysing national trends, and to inform employers, career guidance, education and training services. They include the USA's Occupational Information Network (O*NET) program; the European classification of skills, competencies and occupations (ESCO), and the Australian Core Skills Framework (ACSF). Amongst the strengths of these large-scale classification systems is their breadth in coverage of skills. Nevertheless, they are not particularly detailed or comprehensive in the *definition* of the *individual skills* required to perform a given job role. Consequently, our examination of existing frameworks also encompassed skills and competency models and structures that presented more detailed definitions of individual skills, including the Skills Builder Universal Framework (Ravenscroft and Baker, 2020) and UNICEF's Life Skills and Citizenship Education (LSCE) Framework (UNICEF MENA, 2017).

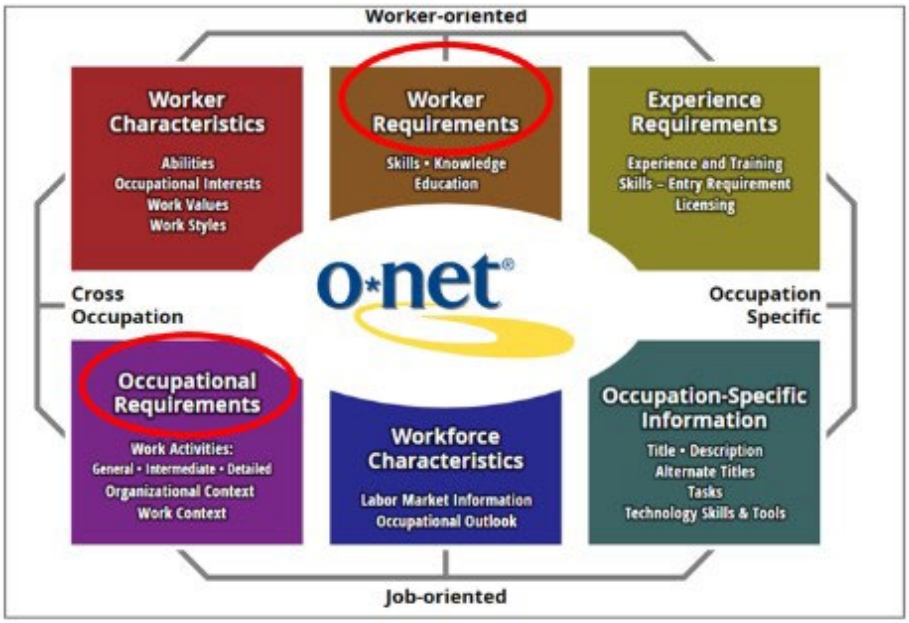
Of the existing large-scale classification frameworks, O*NET was drawn upon most heavily. O*NET had already been used to identify the 'Essential Employment Skills' anticipated to be most intensively utilised across employment in 2035. For a discussion of the merits and limitations of O*NET, we refer readers to our previous paper published as part of The Skills Imperative 2035: *An Analysis of the demand for skills in the labour market* (Dickerson *et al.*, 2023). The conceptual framework underpinning the NFER Essential Employment Skills Survey is based on the O*NET descriptors for each of the six EES[2], combined with the skills definitions from the literature review. This maintains coherence throughout the research project.

The O*NET classification system derives from the O*NET Program and is sponsored by the US Department of Labor/Employment. It is the primary source of descriptive occupational information across the US economy, covering around 1000 occupations (European Commission, 2019). O*NET identifies, defines, describes, and measures the attributes utilised in 964 occupations. Scores for each attribute are based on survey responses from employees and employers, and job evaluation experts observing employees undertaking their jobs (Felstead *et al.*, 2007 and European Commission, DG Employment, Social Affairs and Inclusion, and U.S DOL, 2022). The O*NET content model[3] (see Appendix A of Dickerson *et al.*, 2023) features six major domains, as shown in Figure 1 below. The six EES are located within two of these domains: *Worker requirements*, and *Occupational requirements*.

[2] Note 'Information literacy' combined a set of related O*NET work activities into one EES domain which was considered analogous to the skill 'Critical Thinking'. Hence, the definition, description and measurement of respondents' Skills Supply and Skills Requirement for Information literacy drew on O*NET's conceptualisation and measures for 'Critical Thinking'.

[3] The O*NET® Content Model at O*NET Resource Center (onetcenter.org)

Figure 1 The O*NET content model with annotations highlighting the two areas used to develop our conceptual framework



Source: https://www.onetcenter.org/content.html

Each of the six domains is split into smaller elements containing individual attributes. The '**Occupational requirements'** domain includes: *Generalised work activities* (*information input*, *mental processes*, *work output* and *interacting with others*), *Intermediate work activities*, *Detailed work activities*, *Organisational context*, and *Work context*. The ***Generalised work activities*** area is where five of our six EES are located. Three of our EES are categorised under *mental processes (reasoning and decision making)*: 'Making Decisions and Solving Problems' (i.e. 'Problem solving and decision making' in our set of EES, labelled PSDM), 'Thinking Creatively' (i.e. 'Creative thinking' in our set of EES, labelled TCRE) and 'Organising, Planning and Prioritising Work (i.e. 'Organising, planning and prioritising' in our EES, labelled OPPR). Two of the essential employment skills are categorised under *interacting with others (communicating and interacting)*: 'Communicating with Supervisors, Peers, or Subordinates' (i.e. 'Communication' in our EES, labelled COMM), and 'Establishing and Maintaining Interpersonal Relationships' (i.e. 'Collaboration' in our EES, labelled COCO).

The '**Worker requirements'** domain includes: *Basic skills* (subdivided into *content* and *process*), Cross-functional skills (subdivided into *Social skills* and *Complex problem solving skills*), *Knowledge*, and *Education*. The final of our six essential employment skills ('Information literacy', labelled INLI) drew on O*NET elements related to the appraisal, dissection, synthesis and interpretation of information. As such, it was conceptualised as closely related to the O*NET skill 'Critical Thinking', which falls under ***Basic skills*** and is categorised under *process*. It was also linked to the O*NET skills of: 'Getting Information'; 'Updating and using relevant knowledge'; 'Processing information'; and 'Analysing data or information'.

In order to explain the origins of the main O*NET skill definitions that informed the development of the conceptual model, we present a series of figures (using our own visualisations) that depict the nested locations of the O*NET skills in the O*NET content model, together with their corresponding O*NET definitions.

Figure 2 Location of original O*NET skill 'Establishing and Maintaining Interpersonal Relationships', which informed 'Collaboration' in our EES (our own visualisation)

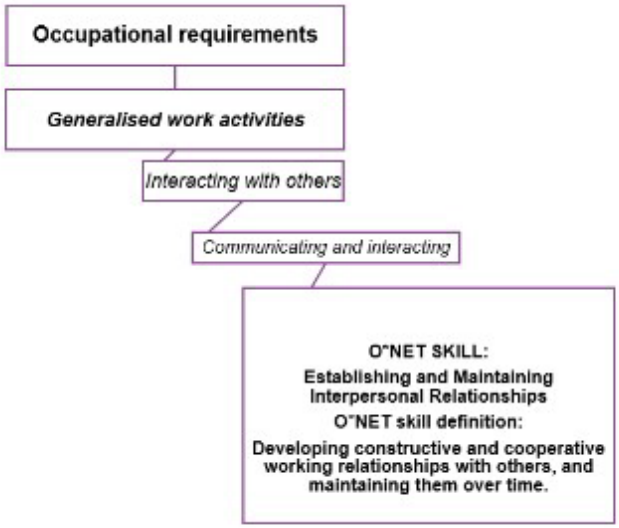


Figure 3 Location of original O*NET skill 'Communicating with Supervisors, Peers, or Subordinates', which informed 'Communication' in our EES (our own visualisation)

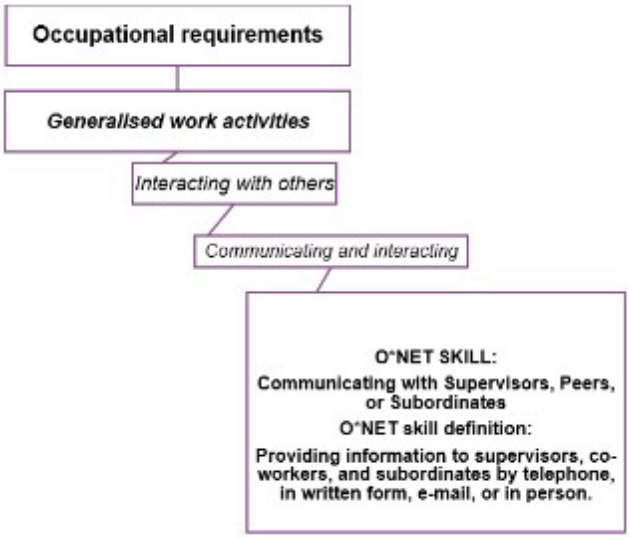


Figure 4 Location of original O*NET skill 'Thinking Creatively', which informed 'Creative thinking' in our EES (our own visualisation)

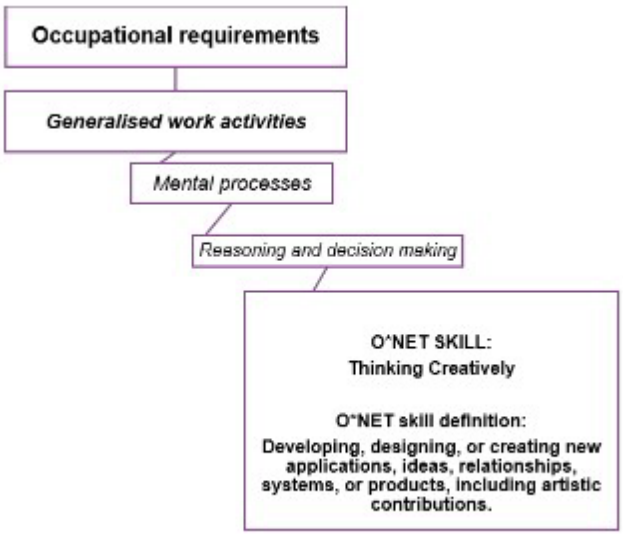


Figure 5 Location of original O*NET skill 'Critical thinking', which informed 'Information literacy' in our EES (our own visualisation)

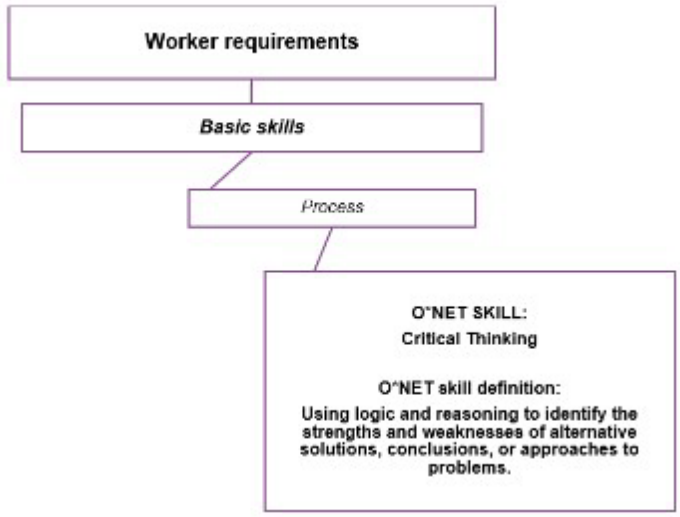


Figure 6 Location of original O*NET skill 'Organising, Planning and Prioritising Work', which informed 'Organising, planning and prioritising' in our EES (our own visualisation)

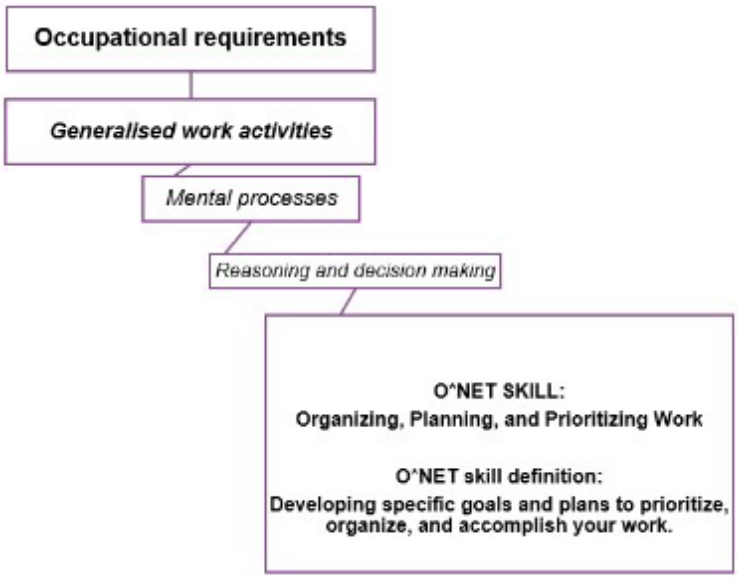
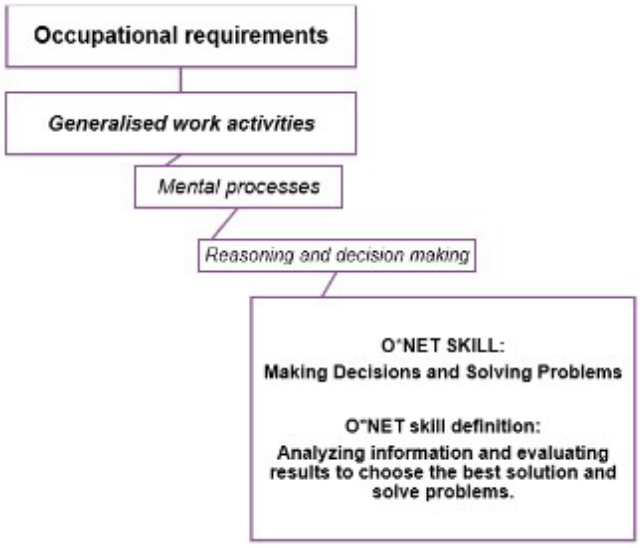


Figure 7 Location of original O*NET skill 'Making decisions and solving problems' which informed ' Problem solving and decision making' in our EES (our own visualisation)



The O*NET classification system is generally accepted to be a robust tool for providing meaningful descriptors of job requirements and is based on 20 years of research and development (European Commission, DG Employment, Social Affairs and Inclusion, and U.S DOL, 2022; Felstead *et al.*, 2007; Popov, Snelson and Baily, 2022). There were some limitations, however, to using the O*NET framework for the development of our own conceptual framework and measures of Skills Supply. The main limitation was that the skills in O*NET are often defined in a very narrow, functional and job-specific manner, which makes some of the elements in the definitions unrelatable to those who are in education,

unemployed or in very different or unusual occupations. The brevity of O*NET skills definitions also makes it difficult to break down each skill into its constituent attributes. O*NET is also slow to adopt new occupations (Popov, Snelson and Baily, 2022).

It was, therefore, necessary to add depth to the definitions and augment them by incorporating other relevant descriptions from different sources. As mentioned above, we drew upon two conceptual skills frameworks that were developed in the context of education. The first framework was the Life Skills and Citizenship Education framework (UNICEF MENA, 2017), developed under the auspices of UNICEF and the World Bank. It has four dimensions (*Learning to know*, *Learning to do*, *Learning be* and *Learning to live together*) and identifies twelve core life skills: *Creativity*, *Critical thinking*, *Problem-solving*, *Cooperation*, *Negotiation*, *Decision-making*, *Self-management*, *Resilience*, *Communication*, *Respect for diversity*, *Empathy*, and *Participation* (Hoskins and Liu, 2019).The second framework, Skills Builder, was developed by a not-for-profit social enterprise originating from the UK, but with international reach. It comprises a universal framework of eight essential skills: *Listening*, *Speaking*, *Problem solving*, *Creativity*, *Staying positive*, *Aiming high*, *Leadership*, and *Teamwork*. As is evident from these brief descriptions, both frameworks identify a number of essential, transversal, 21st century or life skills and have an outlook towards society as well as employment. The advantage of the skills in these frameworks, for development purposes, is that the *definitions* of the skills included are more comprehensive and detailed.

1.2.2 Procedure for creating the thematic areas

The next step in the instrument development process was to break the six EES down into thematic areas. This was done using the following process:

I. Working from the definitions of the specific skill areas from the above-mentioned classification systems and conceptual frameworks, we identified potentially relevant definitions and dimensions within these definitions. A preliminary table containing these definitions was created.

II. The table was refined by removing duplication, and screening for relevance in relation to the O*NET definition.

III. For each skill, the resultant table content was developed into thematic areas that provided content coverage relevant to the O*NET definition. These areas were realised further through an iterative process that involved reviewing the original O*NET definition with respect to relevant definitions and skill descriptions from the other frameworks that had been reviewed. The thematic areas that emerged from this process (presented below) constituted the conceptual framework which served as the basis for item development.

IV. The thematic areas underwent two expert review stages (once before the piloting of the measurement instrument and once afterwards), conducted by Professor Bryony Hoskins. The purpose of the reviews was to check that:

    a. the thematic areas were closely related to the O*NET definitions and made sense

    b. thematic areas were not duplicated across different skills.

V. Responding to the review feedback and pilot data, the skills definitions and areas were adjusted where necessary.

1.2.3 The conceptual framework

In the following section, we present, in tabular form, the thematic areas that were developed for each of the essential employment skills, using the procedure described above. This

comprised the conceptual framework that underpinned the survey development. For each skill area, the tables operationalised the skill as a number of assessable domains. During item writing, each draft item was assigned to a thematic area. This enabled coverage across thematic areas and within assessable domains to be carefully monitored.

Table 1 Thematic areas for Collaboration (COCO)

	Thematic areas
A	Formation of and maintaining constructive/collaborative relationships with others
B	Effective interactions in collaborative situation

Table 2 Thematic areas for Communication (COMM)

	Thematic areas
A	Recognition that communication involves shared meaning
B	Willingness to provide information and understanding about what this involves
C	Adaptation of mode and/or style of delivery in relation to recipient

Table 3 Thematic areas for Creative thinking (TCRE)

	Thematic areas
A	Development of new/different ideas
B	Creation of something new/different
C	Application of a fresh perspective to an issue or challenge
D	Application of thought in a new/different way

Table 4 Thematic areas for Information literacy (INLI)

	Thematic areas
A	Determining appropriate actions using logic and reasoning
B	Identification of strengths and weaknesses through reasoning
C	Evaluation of credibility and reliability of information

Table 5 Thematic areas for Organising, planning and prioritising (OPPR)

	Thematic areas
A	Development of a goal/plan to prioritise something
B	Development of a goal/plan to organise something
C	Development of a goal/plan to complete objectives

Table 6 Thematic areas for Problem solving and decision making (PSDM)

	Thematic areas
A	Analysis of information for problem solving
B	Identification of problems and associated risks and benefits of solutions
C	Using effective strategies for identifying solutions and solving problems
D	Evaluation of information for decision making
E	Using effective strategies for choosing between options

1.3 Item development

1.3.1 Instrument specification

The items for the survey needed to be developed to fulfil the survey instrument specification set out below.

Instrument purpose, scope and format:

- The instrument must measure the availability of EES amongst a representative sample of people aged 16 to 65.
- The instrument must capture the selected six skills identified as Essential Employment Skills (EES): i.e. 1. Collaboration, 2. Communication, 3. Creative thinking, 4. Information literacy, 5. Organising, planning and prioritising, and 6. Problem solving and decision making.
- The instrument must assess the constructs for these skills as defined by the thematic areas (see thematic areas Tables).
- The items will be self-report items, with closed responses.
- The items will be on a six-point Likert-type response scale which would allow for collapsing of response categories as advised by psychometricians.

Sample to be assessed:

- The instrument must be suitable for a representative sample of 16 to 65-year-olds with a wide range of education and qualification backgrounds, currently residing in England.

Accessibility:

- To ensure the instrument is accessible to everyone within the sample (irrespective of education, skills or employment status), the items must be:
    - applicable to situations outside work as well as in work
    - straightforward to access and respond to in terms of cognitive and reading load.

Language:

- The instrument must be written in the English language.

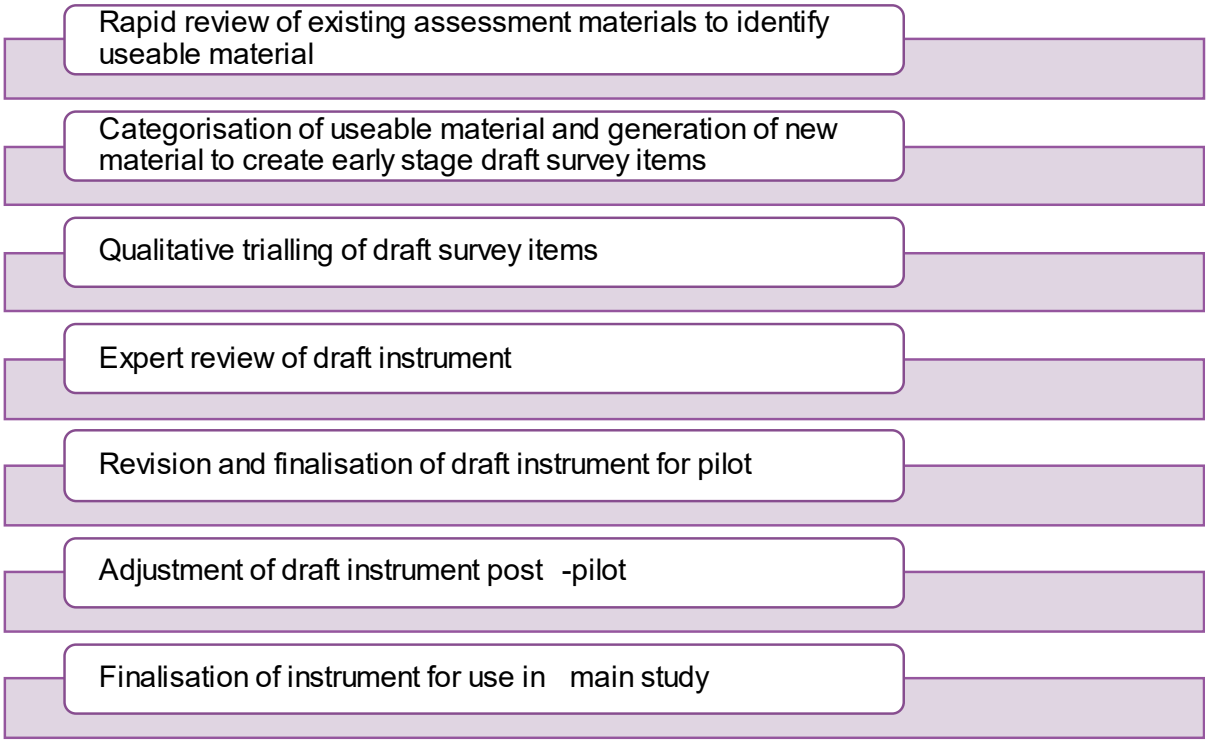
Additional requirements:

The length of the survey was intended to be approximately 8 minutes (discounting the time required to answer background questions), allowing for 72 self-report items in the main study. This translates to a requirement for 12 items per skill area in the main study. To allow for a 25 per cent attrition rate, 16 items per skill area were developed to provide 96 items for the pilot study (see Figure 8).

1.3.2 Instrument development process

In order to develop a robust and valid instrument, the item writing procedure outlined in Figure 8 was adhered to. As presented in Figure 2.1, existing materials in the field were, as a first step, rapidly reviewed to assess their broad suitability in relation to the instrument specification. Once potentially usable materials had been identified, they were categorised to identify which individual items might be utilised (with or without adaptations), and what proportion of new items would need to be generated to meet the specifications for the survey's pilot stage. A draft instrument was designed, always keeping in mind the aim to provide coverage of the thematic areas within a given skill. Interviews were conducted as part of a qualitative trialling process to test out the early-stage draft survey items. This was followed by detailed review of the draft instrument undertaken by subject matter and assessment experts. Each step is now explained in further detail.

Figure 8 Illustration of the item development process



1.3.3 Review of existing measures

NFER conducted a rapid review of existing measures by consulting its own internal database of existing measures, which currently contains over 300 individual entries. Of those, 31 measures were identified as assessing the relevant skill areas and/or constructs. This was narrowed down to five measures that were deemed highly relevant to the project's assessment specification (see Section 2.1). It was necessary for measures/items under consideration to be either freely available to view and use, and/or with permission to use granted. For the five measures identified as extremely relevant, an initial in-depth review was conducted at item level. Each item was considered against the specification and categorised as:

- Potentially usable without adaptation, or
- Potentially usable with adaptation, or
- Not suitable.

It must be noted that the judgment about suitability is entirely based on the relevance to the present instrument specification and does not represent a generic judgement about the inherent quality of the original items and/or instrument.

The rapid review resulted in 39 items deemed potentially suitable for: **Collaboration** (COCO); 43 items potentially suitable for **Communication** (COMM); 15 items potentially suitable for **Creative thinking** (TCRE); 18 items potentially suitable for **Information literacy** (INLI); 9 items potentially suitable for **Organising, planning and prioritising** (OPPR); and 40 items potentially suitable for **Problem solving and decision making** (PSDM).

1.3.4 Developing the draft survey items

From the initial overview, it was evident that, with the exception of the skill Organising, planning and prioritising (EES5), there was a reasonable and, in some cases, good amount of existing material available that could potentially be repurposed for this project's draft survey instrument. During the development process it was, however, borne in mind that existing material may have been created for use with samples and for contexts substantively different from our own (e.g. populations with different age ranges and/or other demographic profiles; clinical contexts; individual one-to-one assessment situations). Items may also have been generated for use with different response scales. Therefore, it was necessary to consider, on an item-by-item basis, whether and what kinds of adaptations and amendments might be necessary to generate items suitable for the given instrument specification. Accordingly, assessment specialists worked on adaptations to the materials where necessary, by changing wording and/or adjusting contexts. In addition, to ensure that the assessment of a given construct provided coverage of the thematic areas in sufficient depth, some new items were generated for all six skill areas to complement the existing materials.

As

Table 7 indicates, 15 to 21 items per skill area were taken forward for qualitative trialling, with the majority originating from existing material. Then, for each skill area, 16 items were selected for the draft pilot instrument. This was reduced following piloting. Thirteen reverse-coded items were included in the initial pilot study. The instrument taken forward to initial piloting had 50 per cent of items on a 6-point frequency scale (Never/almost never/sometimes/often/almost always/always) and 50 per cent on a 6-point agreement scale (Strongly disagree/disagree/slightly disagree/slightly agree/agree/strongly agree).

Table 7 Number of survey items taken forward at each stage of development

Skill	No. items taken forward to qualitative trialling	No. items taken forward to initial pilot survey	No. items taken forward to main stage survey
Collaboration (COCO)	20 (no newly-developed items)	16	12 (2 reverse-coded items)
Communication (COMM)	21 (incl. 5 newly-developed items)	16	12
Creative thinking (TCRE)	20 (incl. 7 newly-developed items)	16	12 (1 reverse-coded item)
Information literacy (INLI)	20 (incl. 3 newly-developed items)	16	12
Organising, planning and prioritising (OPPR)	16 (incl.10 newly-developed items)	16	12 (1 reverse-coded item)
Problem solving and decision making (PSDM)	21 (incl. 2 newly-developed items)	16	12 (2 reverse-coded items)
Total items	118	96	72

	(incl. 27 newly-developed items)		

1.3.5 Qualitative trialling

Qualitative trialling was conducted to test the draft survey items at an early stage in their development. This involved individual interviews with a small number of volunteer participants across England, who were contacted by NFER staff who were already known to them. The sample was not intended to be representative, but coverage of a broad range of ages, locations and occupational experiences was achieved.

To take part in the qualitative trialling, participants had to be between the ages of 18 and 60, live in England and either be currently employed or a university student. Potential participants received an information sheet about the purpose of the qualitative trial and what their involvement would entail. They were asked to provide written consent if they were happy to continue. A total of 16 volunteers meeting these criteria gave their informed consent to participate in the interviews. The participants were all within an age range between 20 to 55 years.

Ahead of the interviews, the six skill areas were split into two sets to make the load manageable for each participant. Participants were accordingly split into two groups, with each participant randomly allocated to a group prior to their interview. One group was assigned items on three of the six EES, which consisted of 61 items; the second group was assigned items from the other three EES, consisting of 56 items. If participants finished their assigned set with time to spare, they were presented with items from the alternative set. As such, each set was seen nine times.

Individual interviews (one participant with one researcher) were conducted by three researchers over the course of a seven-day period, between the 16th and 23rd September 2022. Each interview took between 40 – 60 minutes. During the interview, the researcher shared their screen so that a draft survey item was visible to both parties. As the main goal was to test out the survey questions, participants were firstly asked what response they would give if the question came up in a survey. For this, participants gave their intended response to the Likert-type scale verbally (e.g. *I would never/almost never/sometimes/often/almost always/always do this*; *I strongly disagree/disagree/slightly disagree/slightly agree/agree/strongly agree with this*), which was recorded by the researcher. The participants were then asked to reflect on the item in more detail in terms of whether they considered its meaning to be clear or difficult to interpret. Any feedback about the wording, item difficulty or interpretation was recorded. Once the interview concluded, the responses given by participants for each item were inputted into a spreadsheet to facilitate data analysis.

The researchers considered the feedback from the qualitative trialling and made amendments to items where necessary. These post-qualitative trial revisions involved adjustments to the wording of items and were predicated on participants' interpretations, the perceived difficulty and/or the meaning of the item in relation to their current work context. Following detailed review and further adjustment, a final list of 16 items per skill area was proposed and the draft instrument was finalised for the pilot study.

1.3.6 Review process

A rigorous review process is an integral part of any high-quality item development procedure. At multiple stages, input was requested from experts in the field to support validity. The items for the draft instrument underwent two expert review stages, once before the piloting of the measurement instrument and once afterwards. The review was conducted

by Professor Bryony Hoskins from the University of Roehampton, and specialist assessment researchers at NFER.

The review focused on checking that:

- The items reflected the essence of the O*NET definition and the thematic areas developed for each skill
- The items comprehensively fitted and covered thematic areas identified (at least two items per scale)
- There was no repetition of items across different skill areas
- The items were suitably generic so that they could apply to a very diverse group of people with different work backgrounds, different jobs or no job at all.

Items were then piloted. Instrument piloting is described in ‘Section 2: Pilot and main stage survey design – This describes the design of the pilots and main stage survey. Revisions were then made based on the technical functioning of the items in combination with content expert judgements. This process and the results of the psychometric analysis conducted are discussed in ‘Section 4: Validating and refining the instrument’.

2 Pilot and main stage survey design

2.1 Overview of instrument piloting and main stage survey design

This section describes the design of three survey pilots and the main stage survey design.

1. Pilot 1 (N=3938) – This was the main pilot and was part of the original research design. Its purpose was to test the functioning of the Skills Supply items on a large-scale sample that was representative of the population of 16-65 year olds.
2. Pilot 2 (N=111) – Pilot two was not part of the original research design. Its purpose was to investigate whether the two items (i.e., importance and level) used to measure Skill Requirement in each domain could be placed on the same latent scale as the items measuring Skills Supply. It consisted of a small-scale pilot on a convenience sample and was limited in the number of skill domains and items it included.
3. Pilot 3 (N=83) – Pilot three was also not part of the original research design. Analysis of data from Pilot one suggested that respondents were not using the full rating scale when responding to our Skills Supply items. This might mean our instrument did not fully capture the variability in Skills Supply across the population. The purpose of Pilot 3 was to investigate whether changes to the survey rating scale would result in respondents using the full range of the rating scale. Again, this was done using a small sample of convenience.

The final survey of over 11,000 people is referred to as the 'Main stage survey' (N=11437).

2.2 Pilot 1 design

The purpose of Pilot 1 was to investigate the functioning of the Skills Supply items on a large-scale sample that was representative of the population of 16-65 year olds for whom the items were intended. To reduce response burden, a balanced incomplete block design (BIBD) was used. The BIBD was designed so that respondents were only required to respond to a specific number of item blocks. Each item block was presented to approximately the same number of respondents and piloted in all three positions [i.e., beginning (items 1-24), middle (items 25-48), and end (items 49-72)]. Consequently, the BIBD also catered for a possible order effect, an effect in which survey responses could be affected by the administration item order.

As there were a total of 96 items piloted (16 items × six Skills Supply domains), each item was assigned to one of four blocks per Skills Supply domain. Four pilot instrument versions were created, each containing a total of 72 items (12 items × 6 skills), 12 items (3 blocks × 4 items/block) per skill domain. The content of each of these was largely overlapping, with each variant of the instrument including three of the four blocks of items to measure Skills Supply in each domain. Two batteries of demographic questions were included – these were identical across all four questionnaire variants. Table 8 illustrates the BIBD design. For example, in instrument version 1 (v1), 12 PSDM items (blocks A, B, and C) were administered at the beginning of the instrument, while in instrument v2, 12 PSDM items (blocks B, C, D) were administered at the end of the instrument. In instrument v3, 12 PSDM (blocks C, D, and A) were administered in the middle of the instrument. In instrument v4, 12 PSDM items (blocks B, C, D) were administered at the end of the instrument.

Table 8 Pilot 1 BIBD

Version	Item no.	Instrument v1	Instrument v2	Instrument v3	Instrument v4
Beginning	1 – 12	PSDM (A, B, C)	OPPR (D, A, B)	COCO (C, D, A)	INLI (B, C, D)
	13 – 24	COMM (A, B, C)	TCRE (D, A, B)	OPPR (C, D, A)	TCRE (B, C, D)
Middle	25 – 36	COCO (A, B, C)	COMM (D, A, B)	INLI (C, D, A)	OPPR (B, C, D)
	37 – 48	TCRE (A, B, C)	INLI (D, A, B)	PSDM (C, D, A)	COMM (B, C, D)
End	49 – 60	INLI (A, B, C)	COCO (D, A, B)	TCRE (C, D, A)	PSDM (B, C, D)
	61 - 72	OPPR (A, B, C)	PSDM (D, A, B)	COMM (C, D, A)	COCO (B, C, D)

2.3 Pilot 2 design

The purpose of Pilot 2 was to investigate whether the two items (i.e., importance and level) used to measure respondents' Skills Requirements in each domain could be placed on the same latent scale as the items measuring their Skills Supply in the same domains. The instrument for Pilot 2 consisted of 36 items: 12 OPPR Skills Supply items, 12 PSDM Skills Supply items, and 12 Skills Requirement items (1 importance item + 1 level item × 6 domains). As this pilot was not part of the original research design, a convenience sampling method was used, made up of people contacted by NFER staff who were already known to them. A total of 111 people completed the survey.

2.4 Pilot 3 design

Analysis of responses to Pilot 1 revealed that the majority of items were easy to endorse, implying that respondents were not using much of the bottom part of the Likert scale (i.e., Strongly Disagree, Disagree, and Slightly Disagree). Consequently, another sample of convenience was used to investigate whether replacing the original Likert scale (i.e., 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, and 6 = Strongly Agree) with a frequency scale (i.e., 1 = Never, 2 = Almost Never, 3 = Sometimes, 4 = Often, 5 = Almost Always, 6 = Always) would result in respondents using more of the bottom part of the frequency scale. A total of 83 respondents answered the Pilot 3 survey. Table 9 illustrates the number of items that were piloted with the original Likert scale or the new frequency scale per skill, or both. For three of the six skill domains (i.e., TCRE, OPPR, PSDM), all items used a frequency scale (N), while for the remaining three skills (i.e., COCO, COMM, INLI) utilised both frequence and agreement scales.

Table 9 Pilot 3 Rating scales

Skill domain	Total items	Scale	Likert scale	Frequency scale
COCO	12	Mixed (M)	6	6
COMM	12	Mixed (M)	7	5
TCRE	12	New (N)		12
INLI	12	Mixed (M)	7	5
OPPR	12	New (N)		12
PSDM	12	New (N)		12

2.5 Main stage survey design

The main stage survey consisted of 93 items designed to measure Skills Supply, Skills Requirements and personality traits[4], administered to a total of 11,437 respondents. 72 items measured Skills Supply: 12 in each of the six domains. Three of these domains (COMM, COCO, INLI) used Likert scales eliciting degrees of agreement to the statements presented, whilst the other three domains (PSDM, OPPR, TCE) used frequence scales. Agreement and frequence scales were not mixed within domains. 12 items were used to measure Skills Requirements (asking respondents to rate the 'importance' and 'level' of skill required in each domain). Level and importance were then added together for each domain to create one 'Skill Requirement' score for each domain for each respondent. This is discussed further in 'Section 4: Instrument validation'. 15 additional items were used to measure respondents' levels of the 'Big Five' personality traits.

Tables 10 to 12 report the final number of items in each domain.

Table 10 A breakdown of the number of items used to measure Skills Supply in each domain

	Domain	No. of items	Scale Type
Skills Supply	COCO	12	Likert
	COMM	12	Likert
	TCRE	12	Frequency
	INLI	12	Likert
	OPPR	12	Frequency
	PSDM	12	Frequency
Total		72	

[4] Analysis of respondents' personality traits is intended to feature later in the research programme.

Table 11 A breakdown of the number of items used to measure Skills Requirement in each domain

	Domain	No. of items	Scale Type
Skill requirements	COCO	1	1 – 5
	COMM	1	1 – 5
	TCRE	1	1 – 5
	INLI	1	1 – 5
	OPPR	1	1 – 5
	PSDM	1	1 – 5
Total		6	

Table 12 A breakdown of the number of items per 'Big five' personality trait

	Domain	No. of items	Scale Type
Personality	Extraversion	3	Likert
	Agreeableness	3	Likert
	Conscientious	3	Likert
	Emotional stability*	3	Likert
	Open-Mindedness	3	Likert

* The trait measured was *neuroticism,* but this was reversed to become 'emotional stability' so that all of the personality scales ran in the same direction and could be more easily compared.

3 Data collection and weighting

This section focuses on the approach that was used to collect data for the Pilot 1 survey and the Main Stage survey. Pilots 2 and 3 were small-scale pilots intended to achieve specific objectives and involved testing a limited number of items on a sample of convenience – as such, data collection for these pilots is not discussed further in this section. This section gives an overview of the data collection approach for Pilot 1 and the Main stage survey, before providing further detail on the sampling, fieldwork and weighting specification used for each of these two surveys.

3.1 Pilot 1

3.1.1 Pilot 1 data collection approach

Data was collected using a Computer Assisted Web Interviewing (CAWI) self-completion methodology. Members of research panels convened by Kantar Public were sent invitations to participate in the survey, which included unique links and passwords to access the survey (ensuring that respondents could not complete the survey more than once and ensuring that it was not possible for non-panel members to randomly access the survey). To reduce response burden, a balanced incomplete block design (BIBD) was used (see 'Section 2: Pilot and main stage survey design – This describes the design of the pilots and main stage survey.. Each of the four survey variants was run as a separate survey, with quotas applied and a target sample size of 1,000 for each. These targets were achieved, with a median interview length of between 5 minutes 51 seconds to 6 minutes 2 seconds for each variant.

3.1.2 Pilot 1 Sampling

The sample for Pilot 1 consisted of people in England aged 16-65. The target sample size for each of the four survey variants was n=1,000. Between 1,001 and 1,003 responses were secured for each variant. Survey participants were sourced from pre-recruited double opt-in panels of potential participants run by Kantar Public. Demographic information is available for panel members, which allows the differential sampling of sub-groups within the population according to known response rate patterns. For example, young male panel members are known to be less likely to respond to any survey invitation, so they are deliberately over-sampled in the initial survey invitations.

Quotas were applied for each survey variant to ensure that the achieved sample was representative of the wider population, as shown in Table 13 below. Quota targets were based on February to April 2022 data sourced from the Labour Force Survey (LFS). The quota on education was set up as a monitoring quota. For monitoring quotas, a hard quota is applied during the early stages of fieldwork, to ensure that there are no significant sample skews. Once this has been established, the quotas are then relaxed (i.e. set to a value of 1,000 so that they are effectively deactivated in a sample where n=1,000). All other quotas were relaxed towards the very end of the fieldwork period to avoid the introduction of sample skews. This is to avoid the risk of interviews towards the end of fieldwork being forced to consist of respondents who meet multiple criteria which would not occur frequently in the population as a whole.

3.1.3 Pilot 1 Fieldwork

Potential participants on the research panels were invited to participate via an email with a direct link to the survey with unique login details. Sampling for the invitations was deliberately disproportionate, reflecting the known differential response rates for different sub-groups within the population. Respondents were offered a modest incentive for their participation in the survey. Fieldwork was conducted from the 16th to the 22nd of November 2022. The target of n=1,000 was achieved for each of the four survey variants. Reflecting

the quotas applied, the unweighted sample profile for each survey variant was highly similar, as summarised in Table 13 below.

Table 13 Profile of completed interviews for each survey variant

Quota group	Target % (to 2 dp.)	Representative profile per 1,000 interviews	**Actual** (Instr. v1)	**Actual** (Instr. v2)	**Actual** (Instr. v3)	**Actual** (Instr. v4)
Men 16-34	18.94	190	192	190	190	194
Men 35-54	20.24	203	201	205	204	148
Men 55-65	10.47	105	110	110	114	107
Women 16-34	18.34	184	185	192	184	202
Women 35-54	20.74	208	203	178	196	234
Women 55-65	10.97	110	111	122	111	114
Identify in another way	0.29	-	1	6	3	2
North	27.20	272	273	282	272	287
Midlands	29.70	297	297	306	298	305
South	43.10	431	433	415	432	409
White	82.40	824	846	824	822	808
Other ethnic groups	17.60	176	146	173	165	179
Unspecified	-	-	11	6	15	14
Degree level education	30.60	306	410	424	433	443
Below degree level education	69.40	694	549	547	535	515
Unspecified	-	-	44	32	34	43

All Personally Identifiable Information (PII) was removed at the respondent level. Potentially disclosive variables were also removed. In Pilot 1, only the exact age data was considered to be potentially disclosive - this was accordingly replaced by banded age categories.

3.1.4 Pilot 1 Weighting

The data were weighted to correct small deviations from the quota targets. The variables used in the weighting scheme were the same as those used for the survey quotas. Rim weights were applied. Due to the unweighted profile of each of the four survey variants was very similar, it was possible to apply weights to the combined dataset rather than weighting each of the four variants separately, thereby improving the weighting efficiency. The overall weighting efficiency was 93.5% meaning that the effective sample size across the four survey variants was 3,748. The maximum respondent rim weight was 1.57. The minimum respondent rim weight was 0.59.

The weighting targets were based on February to April 2022 data sourced from the LFS (and shown in the 'Target %' column of Table 13). Given that reliable data on the incidence of non-binary adults in England was not available at the time the data were weighted, this group was left in the proportion which fell out naturally in the Instrument Screening surveys (c. 0.3%).[5] Those who refused to provide their highest level of education were included in the 'below degree level' category. This is because those with lower qualifications are more likely to be reluctant to share this information and the 'below degree level' category was the largest category and therefore the one that respondents are most likely to fall into. Those who did not provide their ethnicity were included in the 'White' category on the basis that this is the largest category and therefore the one that respondents are most likely to fall into. This approach also helps to ensure that the voice of those who are genuinely from minority ethnic groups is not under-represented after weights have been applied.

3.2 Main Stage Survey

3.2.1 Main stage survey data collection approach

All interviews were conducted using a CAWI self-completion methodology. The sample comprised three sub-samples (described below). The questionnaire for each of the three surveys was essentially identical, with only minor differences. The median interview length was 19 minutes. The interview structure alternated the six batteries of questions measuring the Skills Supply and Skills Requirements of each skill in a rating scale format. Background questions were asked in a non-rating scale format. This was done with a view to ensuring that respondent interest levels were maintained throughout the interview (given there is a risk of disengagement when survey participants are asked to complete multiple lengthy batteries of Likert-style questions). The order of the batteries of items related to each skill was randomised from interview to interview to minimise any order effects. The ordering of the non-rating scale questions was consistent in all interviews. The data from each of these three sources was weighted separately. These three weighted samples were then combined, and an overall weight was generated.

3.2.2 Main stage survey sampling

The sample for the final survey comprised three sub-samples. The data from each of these three sources was weighted separately. These three weighted samples were then combined, and an overall weight was generated. The three sub-samples were:

[5] Data from the 2021 census has subsequently been published and shows that 0.5% of the total adult population of England and Wales reported that the gender they identify as is different from their sex registered at birth. The sample profile of the pilot survey does not tie up directly with the Census (being limited to 16-65 year olds) and Census data for this more specific age group is not currently available. Nevertheless, it appears that representation of non-binary adults in the pilot was broadly in line with the wider population.
See:
https://www.ons.gov.uk/news/news/firstcensusestimatesongenderidentityandsexualorientation#:~:text=30%2C000%20(0.06%25)%20identified%20as,in%20a%20different%20gender%20identity

- General population sample of adults 18-65
- Youth 'boost' sample (Year 11 and Year 13s). This was split into:
    - NPD sample
    - ILR sample
- PIAAC recontact sample

3.2.2.1 General population sample of adults 18-65

This comprised a general sample of people in England aged 18-65, drawn from Kantar's Public's random probability Public Voice panel, a general population sample of adults aged 18-65 in England. The sample for the survey was drawn from among the 15,383 respondents to the Public Voice recruitment surveys who were (i) resident in England based on the latest information available, (ii) aged between 18 and 65, (iii) had joined the Public Voice panel, (iv) had not left or been removed from the panel, and (v) were not part of the 'telephone only' subset of the panel. This eligible subset of the Public Voice panel was explicitly stratified by broad age group - stratum 1: aged 18-39 (n=7,252); stratum 2: aged 40-65 (n=8,131). In both strata, every available panel member was sampled for the survey. A subset of stratum 2 (n=1,054) was allocated to a reserve pool and not issued in the first instance. The probability of being allocated to the reserve pool varied with the aim of producing a maximally representative respondent sample from the principal sample issue. In technical terms, stratum 2 panel members were allocated systematically to the reserve pool with a probability proportionate to a size measure ('PPS'). In total, 14,329 panel members were allocated for principal issue, with a further 1,054 allocated to the reserve pool. In the event, the whole of the reserve sample was issued, receiving a mailed letter inviting them to take part in the survey but no other contact.

3.2.2.2 Youth 'boost' sample (Year 11 and Year 13s)

This comprised a survey of young people in Year 11 or Year 13 of state schools and colleges in England, drawn from the National Pupil Database (NPD) and Individualised Learner Record (ILR), with samples systematically drawn from within each stratum after the sample frame was sorted by a range of variables, including SEN provision, prior attainment, Free School Meal eligibility and a range of demographic characteristics. This ensured the sample was representative of the population on these variables.

Based on other similar research projects it was anticipated that it would be necessary to issue 4,500 records into field to achieve the target sample size: 2,250 records for Year 11 and 2,250 records for Year 13, split by database to reflect the population profile (901 from the NPD and 1,349 from the ILR). A reserve sample was also selected from which additional cases could be issued into field should the response rate be lower than expected.

This consisted of the NPD sample and the ILR sample, as detailed below.

NPD sample

The NPD sample frame was explicitly stratified by school year, to allow us to use different sampling fractions for each school year, as shown in

Table 14 below.

Table 14 Population and drawn sample sizes

	Population	Original issue sample	Reserve sample
Year 11	612,656	2,250	2,250
Year 13	214,722	901	901

Within each stratum the sample frame was sorted by a range of variables prior to drawing a systematic sample. This ensured that the selected sample within each stratum was representative of the population by these variables. The sorting variables were:

- Region (former Government Office Regions)
- Gender (male / female)
- Ethnicity (white / asian / black / mixed / other / refused or missing)
- Free School Meal eligibility (yes / no)
- SEN provision (none / SEN without a statement / SEN with a statement)
- KS2 overall performance (quartiles) - Average score for Reading, Maths, and Grammar, Punctuation and Spelling (KS2_READSCORE, KS2_MATSCORE & KS2_GPSSCORE)
- KS4 attainment 8 quartile (Year 13 only)
- Year and Month of birth
- Deprivation (IDACI rank)

As response rates were lower than anticipated, two batches of reserve sample were issued into field. The reserve NPD cases issues into field were selected at random from all reserve cases available. A systematic sample was drawn, with the same stratification variables as were used for the initial sample.

ILR sample

A systematic original issue sample of 1,349 ILR records was selected. A reserve sample of an additional 1,349 cases was also selected. The ILR sample frame was sorted by a range of variables prior to drawing a systematic sample. This ensured that the selected sample was representative of the population by these variables. The sorting variables were:

- Region (former Government Office Regions)
- Gender (male / female)
- Free meal eligibility (Yes / No)
- GCSE maths qualification grade (9, 8, 7, A* or A / 6, 5, 4, B or C / 3, 2, 1, D, E, F or G / U or None / NA / Missing)
- GCSE English qualification grade (9, 8, 7, A* or A / 6, 5, 4, B or C / 3, 2, 1, D, E, F or G / U or None / NA / Missing)
- Deprivation (IDACI rank)

As response rates were lower than anticipated two batches of reserve sample were issued into field. The reserve ILR cases issued into field were selected at random from all the reserve cases available. A systematic sample was drawn, using the same stratification variables as when the initial sample was drawn. In total, 360 reserve ILR cases were issued into field, increasing the total issued ILR sample to 1,709. The first issued batch of reserve

sample consisted of 300 cases. The second issued batch of reserve sample consisted of 60 cases.

3.2.2.3 PIAAC recontact sample

This comprised a recontact survey of people who had previously participated in the Programme for the International Assessment of Adult Skills (PIAAC) between September 2022 and June 2023. This will enable linkage later in the programme between our survey data and the rich employment and skills data collected in PIAAC. Only those PIAAC respondents who gave permission to be recontacted were included in the sample, though recontact agreement rates were high. Of the 4,928 respondents who completed the PIAAC interview, 4,365 (89%) agreed to be recontacted and were issued to the recontact survey. The sample was issued in batches with a view to limiting the gap between completion of the PIAAC survey and completion of the recontact survey.

3.2.3 Main stage survey fieldwork

The survey for all three samples was administered using a self-completion online CAWI method. Sampled individuals were sent invitations to participate in the survey, which included unique links and passwords to access the survey (ensuring that respondents could not complete the survey more than once and that it was not possible for uninvited people to randomly access the survey).

The median interview length was 19 minutes. The mean interview length was 21 minutes (excluding outliers where the length was <1/3 or >3 times the median length). All three versions of the survey were based on essentially identical surveys. The interview was structured to alternate the 6 batteries of Skills Supply & Skills Requirements items with background questions in a non-rating scale format

General population sample of adults 18-65: Sampled members of Kantar's Public Voice panel were sent email invitations to participate in the survey, each containing a unique survey hyperlink. Additional verification was based on panel member's birthdate (including year). SMS messages were also sent to respondents at the same time. Email and SMS reminders were sent to all sampled individuals who did not respond to the initial invitation. Letter reminders were also posted to those who did not respond to initial email and SMS reminders. Including the initial invitation plus reminders, respondents were contacted a maximum of five times.

Youth 'boost' sample (Year 11 and Year 13s): All young people in the Year 11/13 sample were invited to take part in the survey via a letter which provided details about the survey and included a unique survey link and password. Invitations for those in Year 11 were addressed to the parent or guardian and included a letter explaining the survey for the parent or guardian, along with a second letter for the young person. Invitations for those in Year 13 were addressed directly to the sampled young person. Reminder letters were sent to those who had not already responded. Including the initial invitation plus reminders, respondents were contacted a maximum of four times.

PIAAC recontact sample: Those sampled for the PIAAC recontact survey were invited to take part in the survey via a letter which provided details about the survey and included a unique survey link and password. Reminder letters were sent to those who had not already responded. Including the initial invitation plus reminders, respondents were contacted a maximum of four times.

Participants in all three samples were offered of £10 for completing the interview.

Fieldwork was conducted between the following dates:

- General population sample of adults 18-65: 24th May to 18th August 2023

- Youth ‘boost’ sample (Year 11 and Year 13s): 5th May to 18th August 2023
- PIAAC recontact sample: 26th April to 18th August 2023

Table 15 below summarises the volume of sample issue and the number of interviews achieved for each of the three strands of the research programme. It also shows the conversion rate (calculated as the sample issued divided by the number of achieved interviews). The figures quoted are before data cleaning was conducted (see ‘Section: 4 Validating and refining the instrument’).

Table 15 Issued sample, achieved sample and conversion rates

	Issued sample	**Achieved interviews**	**Conversion rate (issued/achieved)**
General population sample of adults 18-65	15,383	7,955	52%
Youth ‘boost’ sample (Year 11 and Year 13s)	5,600	2,046	37%
Year 11 only	2,750	1,047	38%
Year 13 only	2,850	999	35%
PIAAC recontact sample	4,365	2,031	47%

All respondents were provided with links to NFER’s Privacy Notice for the main stage survey:

- General population sample of adults 18-65: www.nfer.ac.uk/skills2035-pv
- Youth ‘boost’ sample (Year 11 and Year 13s): www.nfer.ac.uk/skills2035-1518
- PIAAC recontact sample: www.nfer.ac.uk/skills2035-sas

All data were handled in observance of all relevant data projection legislation including GDPR.

3.2.4 Main stage survey weighting

Variables in the form required for the weighting were derived from responses to the NFER Essential Employment Skills Survey. The variables used were:

- Sex
- Age
- ITL1 region (former Government Office Region)
- Highest qualification level
- Ethnicity
- Long term health condition that limits activity (a lot or a little)
- Economic activity
- Country of birth

Missing information was identified from responses to the PIAAC survey or the Public Voice recruitment survey where possible. Back-filling was not possible for the Youth 'boost' sample. Remaining missing data points were then imputed using a chained equation regression model. This approach was used as it retains the covariance between dimensions. The imputation was done for all three samples simultaneously to ensure consistency.

Data from each of the samples were then weighted using a three-stage process: The data from each of the three samples was weighted separately (Stage 1), before the three weighted samples were combined (Stage 2) and finally an overall weight was generated (Stage 3).

3.2.4.1 Stage 1 of weighting procedure for Main stage survey

General population sample of adults 18-65: The Public Voice sample was weighted to be representative of the general population of England aged 18-65. For every respondent, a *base weight* was calculated that was equal to his/her recruitment survey weight divided by the probability of being sampled for this particular survey. For every respondent, a *propensity score weight* was then estimated, as a function of the recruitment survey variables – this was equal to the estimated odds of being present in the fully weighted recruitment survey dataset. The respondent sample was then weighted (using the raking algorithm) to match the profile of the population. Targets were sourced from the January to March 2023 Labour Force survey and the variables were consistent with those used in the PIAAC weighting. Finally, the final Public Voice weight was trimmed (to a minimum of 0.2 x the mean weight, or a maximum of 5 x the mean weight) to minimise the variance of the weights and ensure that results are not biased by a small number of observations. The sample profile after this trimming was inspected to ensure the sample profile was still a reasonable match to the population profile.

Youth 'boost' sample (Year 11 and Year 13s): The NPD and ILR samples were weighted separately and then combined.

- **NPD sample:** Design weights were generated to compensate for the different sampling fractions that had been used for each school year. A regression model was then used to estimate response probabilities based on sample frame variables. Separate models were used for Year 11 and Year 13. The sample frame variables used were: School year; Region; Ethnicity; IDACI quartile; Gender; Free School Meal eligibility; SEN provision; School phase; KS2 overall performance; KS4 attainment 8 quartile - for Year 13 only; Achieved 5 or more GCSE and equivalents at grades A*-C/9-4 (Level 2) including GCSE English and Maths - for Year 13 only. The estimated response probabilities were inverted to create a non-response weight. The final NPD weight was generated by multiplying the design and non-response weights. The representativity of the sample was then assessed by comparing the final weighted sample profile to population benchmarks. The weighted sample was found to be a good match to the population profile.
- **ILR sample:** All young people had the same sampling probability and as such design weighting was not necessary. A regression model was, however, used to estimate response probabilities based on sample frame variables. The ILR sample frame variables used were: Region; Gender; Free meal eligibility; GCSE maths qualification grade; GCSE English qualification grade; IDACI quartile.

The estimated response probabilities were inverted to create the final ILR weight. The representativity of the sample was then assessed by comparing the final weighted sample profile to population benchmarks. The weighted sample was found to be a good match to the population profile. The NPD and ILR weights were then combined into a single weight scaled to the unweighted sample size.

PIAAC recontact sample: Design weighting was applied to compensate for differences in sampling probability (for the original PIAAC study). The design weights were then used as the input weights for raking. This stage of raking ensured that the sample profile matched the population profile at the margins. The targets for the weighting were sourced from the January to March 2023 Labour Force survey and included Age crossed with sex, ITL1 region (former Government Office Regions), Highest qualification (Level 4+ / below Level 4) crossed with age, Economic Activity, Ethnicity, Country of birth.

3.2.4.2 Stage 2 of weighting procedure for Main stage survey

The weights that were generated separately for each source were then used to generate an overall weight.

First, the PIAAC recontact and General population samples were combined together. The weights from each data source were scaled to be in proportion with the sample size of each sample. Finally, individuals aged 16 or 17 in the PIAAC recontact sample were weighted up to the correct proportion (compensating for the fact that these age cohorts are under-represented by the Public Voice survey). This scaling was based on the January to March 2023 Labour Force Survey.

Next, the NPD and ILR Year 11 and Year 13 boost were combined into the overall dataset. The General population and PIAAC recontact samples already included some young people from these age cohorts. Therefore, when adding in cases from the Youth 'boost' sample, we needed to derive weights that ensured:

- That Year 11 and Year 13 respondents (from all three sample sources) were in the correct proportion relative to the general population
- That within these age cohorts we have the correct level of representation of state educated young people. This is important because the NPD and ILR samples only include state-educated young people, whereas the General population and PIAAC recontact samples also include those that have left education, that are in private education, or that are being home schooled.

The approach we used was to combine the weights as follows:

1. Counts from the NPD and ILR used to draw the Youth 'boost' sample and general population counts from the January-March 2023 Labour Force Survey were used to estimate the proportion of these age cohorts covered by the Youth 'boost' survey. We estimated the proportion in these age cohorts that were:

   a) In state education (and that could therefore have been sampled as part of the Youth 'boost' survey of Year 11 and Year 13s.

   b) Not in state education – this will include those not in education, those in private education, those home schooled, etc. (who could not, therefore, have been sampled as part of the Youth 'boost' survey).

2. The Youth 'boost' samples were then combined into the overall dataset. The weights for all individuals in the Year 11 and Year 13 age cohorts (sampled from all three sources) were scaled to ensure these single year age bands are in the correct proportion relative to the rest of the sample (using the January-March 2023 LFS as the benchmark).

3. Year 11 and Year 13 aged young people in state education from all three sample sources were then identified and the weights of Year 13 respondents were scaled to ensure that they were in the correct proportion. This was not possible for Year 11 students because all Year 11 individuals in the PIAAC recontact and General population samples were in state education.

4. Finally, Year 13 aged young people that were not in state education (sampled via the PIAAC recontact and General population samples) also had their weights scaled to ensure they are in the correct proportion.

3.2.4.3 Stage 3 of weighting procedure for Main stage survey

Following the stages outlined above, a combined base weight for all three sample sources was generated. This was then used as the input for a final stage of raking.

First, all survey respondents were classified into three discrete sub-populations required for the analysis (these are described on the next page). There was also a fourth category that was required: Respondents that could not be classified due to providing at least one non-informative response to a question used in the derivation of these sub-populations[1].

Finally, the sample included a small number of respondents (23) in Y12 of education. It was decided to exclude these individuals from the weighting and analysis, as they would need to be upweighted significantly and this would greatly reduce the precision of estimates generated. As such, weighted responses should be deemed to be representative of Year 11 pupils, Year 13 pupils, and the general population aged 18+.

The combined base weight was then used for a final stage of raking to ensure the overall sample profile matched the population profile at the margins. The targets for the weighting were sourced from the January to March 2023 Labour Force survey. All demographic variables included in the weighting were nested within the sub-populations required for analysis. The targets for each variable and category used in the final stage of raking are presented in Appendix A

3.2.4.4 Design effects due to the final weighting

The design effect due to the final weighting[2] has been estimated overall and for the three sub-populations (see next section for an explanation of these sub-populations). The effective sample size figures presented below are for the overall analysis. Questions routed to a subset of respondents have smaller effective sample sizes and wider margins of error.

Table 16 Design effects and effective sample sizes

	Estimated design effect	Estimated effective sample size
Overall	2.47	c.4,638
Workers	1.93	c.4,460
Young people	5.02[6]	c.376
Long-term unemployed	2.53	c.260

[6] This design effect is large due to the upweighting of individuals in this age cohort that were not part of the NPD and ILR boost sample. Specifically, those not in state education, and 18 year olds that are in the cohort above those in Year 13 in 2022/23.

3.2.5 Categorising weighted respondents into one of three sub-populations for analysis

For analysis, we classify respondents into three[4] discrete sub-populations:

1. ***Workers*** **(N = 8,569)**: Adults aged 19-65 who are either currently in paid work or who have been in work at any point in the previous five years, and young people aged 16-18 who are in work-based training or employment 20+ hours per week.
2. ***Young people*** **(N = 1,889)**: 15-18 year olds who are not in work or who are working less than 20 hours per week[7].
3. ***Long-term unemployed*** **(N = 649)**: Adults aged 19 or over that have never worked or who have been unemployed for 5 of more years.

We examine Skills Supply within each sub-population and compare between sub-populations. Analysis of Skills Requirements and Skills Gaps is limited to the 'Workers' subpopulation; for obvious reasons, young people yet to enter the labour market and the long-term unemployed are not asked to rate the skills requirements of their current or most recent job.

[7] We assume that young people working under 20 hours p/w are involved in 'causal work' alongside their studies, whereas those working over 20 hours p/w are involved in work-based learning such as an apprenticeship or have left education to enter the labour market.

4 Validating and refining the instrument

This section describes the analysis of data from all three pilots and the Main Stage survey using the Rasch model. First, Rasch methodology for analysis is briefly described. The Rasch analysis of each pilot is the described and the results reported, and a summary is provided of the changes that were made to the survey after each pilot. Finally, the Rasch analysis of the Main stage survey is described and the results are reported.

4.1 Rasch Methodology

Rasch measurement theory (RMT) was used to enable valid comparisons of Skills Supply between people, and of individuals' Skills Supply and Skills Requirements (enabling 'Skills Gaps' to be calculated for respondents in the 'Workers' subpopulation). RMT provides a robust, flexible, and theoretically sound framework for quantifying latent traits on a linear scale using observed responses to survey items. Rasch analysis allows for the calibration of items that capture both simple and complex constructs, ensuring that each item contributes meaningfully to the measurement of the underlying trait. Using Rasch techniques enables people's raw scores to be expressed on a logit (interval) scale and a 'difficulty' measure to be computed for each item on the same logit scale. Our resulting measures for Skills Supply and Skills Requirements are comparable between people and across items because they account for both individual differences in skills and differences between items in how easy / hard it is to select frequency or agreement responses at the top of the rating scale.

The Rasch measurement model provides a method for assessing measurement invariance, essential for comparing responses between groups or conditions without bias. The property of invariance ensures that the measurement of essential employment skills is consistent across different samples of persons and sets of items. The Rasch model assumes that item responses are independent of each other, meaning that the probability of endorsing one item does not depend on the responses to other items. Both item measures (i.e., the difficulty of each item) and person measures (i.e., the ability of each respondent) are placed on the same latent trait continuum, allowing for a direct comparison between an individual's ability and the difficulty of the items they respond to. It also provides insights into individual performance relative to the items.

The Rasch measurement model also provides fit statistics for both items and persons. Fit statistics indicate how well each item and/or person fits the Rasch model. Items not fitting the model may indicate issues such as poor item wording or construct irrelevant variance, while persons not fitting the model may indicate response patterns that deviate from the expected model. Both item and person fit statistics can be further investigated for validity and reliability of the survey instrument.

Overall, the rigorous framework of Rasch analysis helps validate the constructs in our survey of Essential Employment Skills by ensuring that items align well with the underlying theoretical model. The Rasch analysis of data from each pilot and the Main stage survey are nor described and the results reported.

4.2 Pilot 1

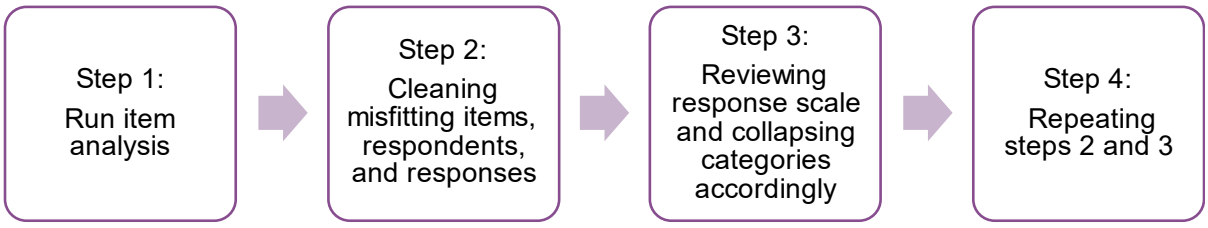
4.2.1 Pilot 1 initial data cleaning

Initial analysis revealed that seven of the 4009 respondents had no data and a further 35 respondents did not have useable data in our six EES domains and hence had to be dropped, resulting in a dataset containing 3967 respondents. The 13 items that were negatively worded (e.g., “I dislike it when people challenge my views”) were also reverse coded for analysis.

The data were analysed through Rasch measurement theory (RMT) using Winsteps® software (version 5.4.1.0, Linacre, 2023a). The original analysis was performed by employing the Rating Scale Model (RSM), a model in which items share the same scale. However, as more items and/or scales were added to the main stage analyses, pilot 1 data were reanalysed using a Partial Credit Model (PCM), a model in which each item has its own scale. The PCM is recommended when analysing survey data with several subscales as a RSM may result in the data becoming multidimensional in such occasions (Linacre, 2023b). Both models flagged the same items for review and, as such, what follows is the reanalyses of Pilot 1 data using PCM.

The data were cleaned in a 4-step process in accordance with Rasch measurement theory (RMT). Step 1 entailed running the analysis with all items (n = 96), step 2 involved dropping items, respondents, and/or responses exhibiting misfit. Item misfit occurs when respondents answer items in an unpredictable way, not as expected by Rasch model. Outfit mean-square values (MNSQ) values outside defined thresholds indicate outliers in the data, while infit MNSQ values greater than defined thresholds are a threat to validity. Employing Linacre’s criteria for survey data (Linacre, 2023b), infit and outfit MNSQ values greater than 1.4 (MNSQ), with a standardized z-score (Zstd) ≥ 2 were flagged as misfit. However, prior to deleting any misfitting items, a procedure suggested by Wright and Stone (1979) was applied. The procedure entails removing individual responses to misfitting items to bring MNSQ values into an acceptable range. If the procedure still resulted in misfitting items, those items were dropped from the analysis. Respondents that had MNSQ values greater or equal to 5 (MNSQ ≥ 5, Wright and Stone, 1979) were dropped from the analysis. Step 3 entailed investigating the original 6-band rating scale structure (i.e., 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, and 6 = Strongly Agree) by applying Linacre’s rating scale guidelines (Linacre, 2004). When the rating scale did not function appropriately, adjacent categories (where applicable) were collapsed. For example, “Strongly Disagree with Disagree” or “Strongly Disagree, Disagree, and Slightly Disagree” were collapsed. Step 4 consisted of repeating the process until there were no misfitting items or overly misfitting respondents (i.e., MNSQ ≥ 5.0) and all items had a well-functioning scale. Figure 9 below illustrates the four-step Rasch cleaning process.

Figure 10 The 4-step Rasch cleaning process



The Rasch cleaning process resulted in 13 items (13.5%) and a further 29 respondents being dropped from any further analysis as they were deemed misfitting. Consequently 83 items (86.5%) and 3938 respondents were retained for analysis.

4.2.2 Pilot 1 analyses (83 items)

The first step in the analysis was to confirm that the cleaned dataset met RMT assumptions. The assumptions of unidimensionality and local item independence were investigated by conducting Principal Component Analysis of Residuals (PCAR) and by comparing standardized residual correlations, respectively. Unidimensionality was evaluated by examining (1) the unexplained variance in the first contrast and (2) the disattenuated correlation between (i) person measures on items in Cluster 1 and person measures on items in Cluster 2 and (ii) person measures on items in Cluster 2 and person measures on items in Cluster 3. Local item dependency (LID) was assesssed by examining the standardized residuals between pairs of items as correlations around 0.70 imply that LID may exist (Linacre, 2023b).

The PCAR analysis revealed that no issues of dimensionality or local item dependency were detected, implying that the data were unidimensional, measuring only one construct, and no pairs of items were locally dependent. Table 17 displays the summary of the Pilot 1 instrument's psychometric properties. Both indices from Rasch Measurement Theory (RMT) and Classical Test Theory (CTT) are presented. RMT measures are expressed in logits (i.e., a Rasch unit of measurement), while CTT indices are expressed in, or based on, raw scores. Higher measures imply it is easier to endorse (agree with) an item.

Table 17 Pilot 1 summary of psychometric properties

No. of respondents	3938
No. of items	83
Mean measure (Mean score)	0.42 (278.8)
S.D. population (S.D. score)	1.02 (44.6)
Max. measure (Max. score)	6.95 (378.0)
Min. measure (Min. score)	–6.71 (62.0)
Person reliability [Alpha (rxx)]	0.95 (0.93)
Person RMSE (SEM)	0.23 (11.59*)
Person separation	4.35
Person strata	6.13
Item measure (RMSE)	0.00 (0.03)
Item reliability	0.99
Item Infit MNSQ (Zstd)	1.22 (8.70)
Item Outfit MNSQ (Zstd)	1.28 (9.90)

The instrument had a Rasch person reliability of 0.95 with an associated person separation of 4.35 and a person strata of 6.13, implying that the instrument was able to statistically distinguish at least six distinct respondent ability groups. Person strata rather than person separation is used to indicate the number of respondent endorsement levels an instrument can statistically discriminate when very low respondent scores and very high respondent scores are truly representative of respondent endorsement. The instrument also had an item reliability of 0.99, the minimum threshold being 0.90, indicating that the sample size of the instrument was large enough to confirm its construct validity. The instrument's corresponding classical test theory reliability (i.e., Cronbach's Alpha) was 0.93.

4.2.4 Pilot 1 analyses by domain

The next step in the analysis was to investigate Skills Supply by domain. For each domain, both the original (Pre) and final (Post) indices are presented. It should be noted that the measures are not comparable across the six domains as each domain was analysed on its own. Table 18 reports the original and final Skills Supply indices.

Table 18 Pilot 1 summary of original and final skill supply domain indices

Skill	Status	No. of items	Person Reliability (RSME)	Alpha (SEM*)	Item Infit (Zstd) [max]	Item Outfit (Zstd) [max]	Item reliability
COCO	Pre	16	0.77 (0.46)	0.73 (4.89)	1.69 (9.90)	1.83 (9.90)	1.00
COCO	Post	13	0.82 (0.67)	0.76 (4.20)	1.32 (9.90)	1.36 (9.90)	1.00
COMM	Pre	16	0.78 (0.49)	0.72 (5.35)	1.80 (9.90)	2.31 (9.90)	0.99
COMM	Post	15	0.82 (0.54)	0.73 (5.09)	1.13 (5.01)	1.16 (5.98)	0.99
TCRE	Pre	16	0.78 (0.46)	0.75 (5.40)	2.23 (9.90)	2.39 (9.90)	1.00
TCRE	Post	14	0.86 (0.62)	0.79 (4.97)	1.26 (9.21)	1.26 (9.10)	0.99
INLI	Pre	16	0.79 (0.53)	0.76 (4.59)	1.95 (9.90)	2.29 (9.90)	0.99
INLI	Post	14	0.83 (0.65)	0.79 (4.04)	1.21 (7.89)	1.22 (8.08)	0.99
OPPR	Pre	16	0.76 (0.44)	0.74 (5.39)	2.13 (9.90)	2.39 (9.90)	1.00
OPPR	Post	14	0.84 (0.59)	0.79 (4.90)	1.20 (7.88)	1.23 (8.04)	0.99
PSDM	Pre	16	0.74 (0.45)	0.71 (4.94)	1.68 (9.90)	1.84 (9.90)	0.99
PSDM	Post	13	0.84 (0.67)	0.77 (4.16)	1.33 (9.90)	1.39 (9.90)	0.99

* approximate due to missing data by design

The original (Pre) person reliabilities for all six skills ranged from 0.74 to 0.79 while maximum item infit and outfit measures ranged from 1.68 to 2.23 and 1.83 to 2.39, respectively. The post cleaning reliabilities ranged from 0.82 to 0.86, while maximum item infit and outfit measures ranged from 1.13 to 1.33 and 1.16 to 1.39, respectively. This suggests the Rasch cleaning process was effective.

4.2.5 Pilot 1 final selection of items (71 items)

The Pilot 1 instrument consisted of more Skills Supply items (N = 96) than were planned to be administered in the main stage survey. Assessment specialists reviewed the analysis and selected the items to retain. This was based on how relevant an individual item was for the assessment of a given skill area, and how well it reflected the underlying thematic areas (i.e. a roughly balanced reflection of items within each of the thematic areas). This judgement was based on the expert review process outlined in 'Section One: Instrument Development'. Items that showed high overlap or statistical correlations with similar items were removed based on an individual item-by-item review. As such, the content experts made the final decision about which items to select for the final main study. Once items were selected, a final check was made to ensure that the final items reflected the feedback from the qualitative trialling and pilot data. Skills definitions, thematic areas and items were adjusted as necessary. At this stage, a total of 25 items, including the 13 misfitting items, were dropped from the main stage survey, resulting in 71 items being kept, 5 domains consisting of 12 Skills Supply items and one domain (TCRE) consisting of 11 items.

To investigate how the remaining 71 items would behave in the main survey, the 4-step cleaning process was repeated. The PCAR analysis confirmed that there were no issues with dimensionality or local item dependency. Table 19 Pilot 1 final selection of items summary of indices (71 items)reports the summary of the final 71-item instrument's psychometric properties.

Table 19 Pilot 1 final selection of items summary of indices (71 items)

No. of respondents	**3938**
No. of items	71
Mean measure (Mean score)	.43 (239.8)
S.D. population (S.D. score)	1.05 (38.8)
Max. measure (Max. score)	6.84 (330.0)
Min. measure (Min. score)	-6.57 (51.0)
Person reliability [Alpha (rxx)]	0.94 (0.93)
Person RMSE (SEM)	0.25 (10.55*)
Person separation	4.12
Person strata	5.83
Item measure (RMSE)	0.00(0.03)
Item reliability	0.99
Item Infit MNSQ (Zstd)	1.23 (9.24)
Item Outfit MNSQ (Zstd)	1.32 (9.90)

* approximate due to missing data by design

The instrument had a person reliability of 0.94, with a person separation index of 4.12 and a person strata of 5.83, implying that the instrument separated respondents into at least 5 categories. The instrument's corresponding classical test theory reliability (i.e., Cronbach's Alpha) was 0.93. The assumptions of unidimensionality and local item independence were met. Table 20 reports the indices for the final selection of items to measure Skills Supply in each domain.

Table 20 Pilot 1 Summary of final skill supply domain indices

Domain	Status	No. of items	Person Reliability (RSME)	Alpha (SEM*)	Item Infit (Zstd) [max]	Item Outfit (Zstd) [max]	Item reliability
COCO	Final	12	0.81 (0.70)	0.77 (4.02)	1.29 (9.90)	1.34 (9.90)	1.00
COMM	Final	12	0.78 (0.61)	0.71 (4.51)	1.12 (4.80)	1.15 (5.63)	.99
TCRE	Final	11	0.84 (0.72)	0.76 (4.20)	1.27 (9.90)	1.35 (9.90)	.98
INLI	Final	12	0.81 (0.70)	0.84 (3.55)	1.17 (6.05)	1.21 (7.00)	.99
OPPR	Final	12	0.82 (0.64)	0.82 (4.30)	1.17 (6.69)	1.20 (6.96)	.99
PSDM	Final	12	0.83 (0.72)	0.80 (3.83)	1.27 (9.76)	1.29 (9.90)	.99

The Rasch person measure reliabilities for all skill supply domains ranged from 0.78 (COMM) to .84 (TCRE), implying that the minimum reliability threshold of .70 had been reached. The item reliabilities for all six domains were above .90, the minimum threshold, indicating that the sample size per domain was large enough to confirm its construct validity. The final Skills Supply domain reliabilities had classical test theory reliabilities (i.e., Cronbach's Alpha) ranging from 0.71 to 0.84. No items were identified as misfitting as all misfitting items had been dealt with in the 4-step Rasch cleaning process.

4.3 Pilot 2

The purpose of pilot 2 was purely to test whether the two items (i.e., importance and level) used to measure Skill Requirement in each domain could be placed on the same latent scale as the items measuring Skills Supply. It consisted of a small-scale pilot on a sample of convenience.

4.3.1 Pilot 2 data cleaning and recoding

Pilot 2 data was again analysed through Rasch measurement theory (RMT). The Rating Scale Model (RSM) was used for this analysis as there were not enough responses per category (i.e., 10 observations per category) for the Partial Credit Model (PCM) to be used. The same 4-step Rasch cleaning process described earlier was used. The two items (i.e., importance and level) measuring Skills Requirements were combined within each domain and then rescored as follows:

‘Importance’ was measured on a scale of 1 through 5, while ‘Level’ was measured on a scale of 1 through 7. If a respondent chose “1” for the first item, their response for both items was treated as missing. For each domain, both items were added to create one item on a raw scale ranging from 3 to 12, which was then converted into a 5-band scale as follows: 3 – 4 raw scores = 1; 5 – 6 raw scores = 2; 7 – 8 raw scores = 3; 9 – 10 raw scores = 4; 11 – 12 raw scores = 5.

4.3.2 Pilot 2 analyses (30 items)

The Pilot 2 instrument consisted of 30 items (12 OPPR Skills Supply items, 12 PSDM Skills Supply items, and 6 Skills Requirements items). The PCAR analysis confirmed that the data was unidimensional as the disattenuated correlation between clusters 1 and 2, and clusters 2 and 3 were at least 0.71, the threshold for both clusters to be considered containing more items measuring the same construct than containing more items measuring different constructs. No issues of local dependency were detected. Table 21 displays the summary of the instrument’s psychometric properties.

Table 21 Pilot 2 Summary of psychometric properties (30 items)

No. of respondents	111
No. of items	30
Mean measure (Mean score)	0.95 (138.9)
S.D. population (S.D. score)	0.93 (16.2)
Max. measure (Max. score)	3.71 (168.0)
Min. measure (Min. score)	-1.06 (90.0)
Person reliability [Alpha (r_{xx})]	0.90 (0.89)
Person RMSE (SEM)	0.30 (5.49*)
Person separation	2.95
Person strata	4.27
Item measure (RMSE)	0.14
Item reliability	0.94
Item Infit MNSQ (Zstd)	1.38 (2.80)
Item Outfit MNSQ (Zstd)	1.39 (2.75)

* approximate due to missing data

The instrument had a person reliability of 0.90, with a person separation index of 2.95 and a person strata of 4.27 implying that the instrument separated respondents into at least 4 categories. Table 22 reports the indices for the two skill supply domains and the skill requirements items. The instrument had a classical test theory reliability (i.e., Cronbach's Alpha) of 0.89.

Table 22 Pilot 2 domain indices

Domain	Status	No. of items	Person Reliability (RSME)	Alpha (SEM*)	Item Infit (Zstd) [max]	Item Outfit (Zstd) [max]	Item reliability
OPPR	FINAL	12	.81 (.57)	0.79 (3.93)	1.25 (1.86)	1.26 (1.76)	0.95
PSDM	FINAL	12	.86 (.87)	0.93 (2.40)	1.34 (2.21)	1.34 (2.14)	0.94
Skill requirements	FINAL	6	.85 (.78)	.87 (1.58)	1.31 (2.08)	1.31 (2.07)	0.95

Both skills supply domains and the skill requirements items exhibited appropriate psychometric indices as person reliabilities ranged from 0.81 to 0.86 and fit statistics ranged from 1.25 to 1.34. As such, the Pilot 2 analysis succeeded in confirmed that the two items measuring skill requirements per domain can be combined, rescaled (see section 3.2 for description) and included in a survey measuring respondents Skills Requirements as well as Skills Supply. The classical test theory reliabilities (i.e., Cronbach's Alpha) ranged from 0.79 to 0.93.

4.4 Pilot 3

Following the analysis of Pilot 1 data, revisions were made to the survey design with the aim of making the item sets more difficult to endorse and to more fully capture the variability in skill levels in the population. The response scales were adjusted for all six skill areas (i.e. all 72 items) so that they focused on the *frequency* (Never/almost never/sometimes/often/almost always/always) with which people may exhibit the behaviours in the statements. As a result, it was necessary to undertake minor wording changes to items to signal frequency responses were required.

4.4.1 Pilot 3 data cleaning and recoding

Once again, Pilot 3 data was analysed through Rasch measurement theory (RMT) using Winsteps® software (version 5.4.1.0, Linacre, 2023a). The Rating Scale Model (RSM) was used for the analysis as there were not enough respondents for the Partial Credit Model (PCM) to be used. The same 4-step process for cleaning the data (see Section 2.3 for description) was used. The two items (i.e., importance and level) measuring Skills Requirements in each domain were combined in the same way as for the Pilot 2 analysis (see Section 3.2 for description). Consequently, the Pilot 3 instrument consisted of 78 items (6 domains × 12 Skills Supply items + 6 domains x 1 Skills Requirement score).

4.4.2 Pilot 3 analyses (78 items)

The PCAR analysis confirmed that the data was not unidimensional and that were issues of local item dependency. Consequently, the analysis that follows is a separate analysis of Skills Supply and Skills Requirements for each of the six EES domains. Table 23 reports the final indices for the Pilot 3 instrument.

Table 23 Pilot 3 skills supply subdomain and skill requirements indices

Domain	Scale*	No. of items	Person Reliability (RSME)	Alpha (SEM**)	Item Infit (Zstd) [max]	Item Outfit (Zstd) [max]	Item reliability
COCO	M	12	0.81 (0.62)	0.84 (2.13)	1.31 (1.96)	1.31 (1.68)	0.99
COMM	M	12	0.75 (0.50)	0.73 (2.84)	1.27 (1.70)	1.26 (1.58)	0.92
TCRE	N	12	0.90 (0.56)	0.92 (2.39)	1.33 (1.91)	1.35 (1.91)	0.95
INLI	M	12	0.82 (0.65)	0.86 (2.09)	1.32 (1.88)	1.31 (1.88)	0.93
OPPR	N	12	0.73 (0.39)	0.77 (2.96)	1.26 (1.74)	1.27 (1.76)	0.99
PSDM	N	12	0.86 (0.54)	0.92 (2.22)	1.38 (2.32)	1.36 (2.10)	1.00
Skill requirements	M	6	0.84 (0.86)	0.96 (0.98)	1.27 (1.58)	1.29 (1.56)	0.97

* M = mixed scale (i.e., Likert and frequency scales) and N = new scale (i.e., frequency scale only)

** approximate due to missing data

The psychometric properties of the Skills Supply scales from Pilot 3 for three domains (TCRE, OPPR, PSDM) were assessed. These were the domains in which rating scales had been switched from Likert-scales soliciting degrees of agreement to frequency scales. This increased reliability in two of the domains (TCRE, PSDM), although reliability in the third domain decreased (OPPR). Importantly, the reliability of the Skills Requirements items remained high across Pilots 2 and 3. The person reliabilities for Pilot 3 ranged from 0.73 to 0.90, while their classical test theory reliabilities (i.e., Cronbach's Alpha) ranged from 0.73 to 0.96.

4.5 Main stage survey

4.5.1 Main stage survey data cleaning and recoding

The Partial Credit Model (PCM), in which each item has its own scale, was used for analysis. The 4-step cleaning process resulted in 15 Skills Supply items (approx 21% of the Skills Supply items) being dropped from the instrument. Table 24 illustrates the domains from which the 15 Skills Supply items were dropped. No items were dropped from the OPPR scale, only 1 item was dropped from PSDM and COMM, but 3 items were dropped from INLI, 4 from COCO and 6 from TCRE. This was necessary to meet the assumptions of unidimensionality and local item independence. PCAR analysis detected no issues with these assumptions in the remaining 78 items.

Table 24 A breakdown of the final number of Skills Supply items by domain

	Domain	Original No. of items	Final No. of items	Scale Type
Skills Supply	COCO	12	8	Likert
	COMM	12	11	Likert
	TCRE	12	6	Frequency
	INLI	12	9	Likert
	OPPR	12	12	Frequency
	PSDM	12	11	Frequency
Total		72	57	

4.5.2 The Main stage survey analyses

Data from the Main stage survey was weighted to be representative of the population and each of the three subpopulations, as detailed in ‘Section 3: Data collection and weighting’. To put Skills Supply and Skills Requirements for each domain on the same latent scale, all items and rating scale thresholds were anchored to their respective difficulty measures retrieved from the analysis of the 78 Skills Supply and Skills Requirements items included in the Main stage survey. Consequently, direct comparisons between Skills Supply and Skills Requirements, and of Skills Supply across domains and groups, can be performed. Table 25 through to Table 27 report the psychometric properties of the Skills Supply items and Skills Requirement items by domain. Overall, the Main stage survey analyses revealed that the Rasch measures were robust and reliable, and as such, valid inferences could be ascertained from them. See Appendix B Main study Item indices through E for Rasch measures, item descriptions, item categories, and item thresholds.

Table 25 Main stage survey psychometric properties of Skills Supply scales by domain

Properties	Skills Supply	COCO	COMM	TCRE	INLI	OPPR	PSDM
No. of (weighted) respondents	11249.8	11249.4	11244.7	11249.5	11238.3	11248.9	11249.8
No. of items	57	8	11	6	9	12	11
Mean measure (Mean score)	0.91 (176.0)	0.92 (23.6)	0.95 (30.8)	0.90 (17.4)	1.00 (26.4)	0.99 (40.2)	1.04 (37.7)
S.D. population (S.D. score)	1.13 (28.7)	1.32 (4.7)	1.35 (6.0)	1.40 (4.9)	1.56 (5.1)	1.34 (9.5)	1.51 (8.3)
Max. measure (Max. score)	7.41 (246.0)	5.88 (35.0)	5.56 (44.0)	5.32 (29.0)	5.13 (36.0)	5.59 (60.0)	6.21 (55.0)
Min. measure (Min. score)	-7.75 (37.0)	-6.52 (5.0)	-5.77 (5.0)	-5.39 (3.0)	-6.63 (4.0)	-5.25 (3.0)	-5.86 (5.0)
Person reliability [Alpha (r_{xx})]	0.95 (0.95)	0.75 (0.86)	0.79 (0.91)	0.76 (0.84)	0.77 (0.94)	0.82 (0.90)	0.84 (0.91)
Person RMSE (SEM)	0.26 (6.09*)	0.66 (1.77*)	0.62 (1.77*)	0.68 (1.93*)	0.74 (1.29*)	0.57 (3.05*)	0.61 (2.47*)
Person separation	4.20	1.72	1.95	1.80	1.85	2.14	2.26
Person strata	5.93	2.63	2.93	2.73	2.80	3.19	3.35
Item mean measure (RMSE)	-0.11 (0.01)	-0.28 (0.02)	-0.16 (0.02)	0.36 (0.01)	-0.63 (0.02)	0.16 (0.01)	-0.08 (0.01)
Item reliability	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Item Infit MNSQ (Zstd)	1.26 (9.90)	1.00 (-0.14)	0.93 (-5.50)	0.87 (-7.66)	0.87 (-9.90)	1.03 (1.90)	0.94 (-4.22)
Item Outfit MNSQ (Zstd)	1.25 (9.90)	1.05 (3.42)	0.98 (-1.65)	0.87 (-7.41)	0.87 (-9.16)	1.02 (1.38)	0.92 (-5.29)

* approximate due to missing data

Overall, the Skills Supply scales had a person reliability of 0.95, with a person separation of 4.20 and person strata of 5.93, implying that the scales were sensitive enough to distinguish amongst approximately 6 levels of endorsement. For each domain, reliabilities ranged from 0.75 to 0.84, with person separations ranging from 1.72 to 2.26, and person strata ranging

from 2.63 to 3.35, implying each domain was sensitive enough to distinguish amongst at least 2 (e.g., disagree and agree) to 3 (e.g., never, sometimes, always) levels of endorsement. Corresponding classical test theory reliabilities (i.e., Cronbach's Alpha) ranged from 0.84 to 0.94. On average, respondents found the PSDM domain (Mean measure = 1.04) easier to endorse and found the COCO domain (Mean measure = 0.92) harder to endorse.

Table 26 Main stage survey psychometric properties for Skills Requirements by domain

Properties	Skills Requirements	COCO	COMM	TCRE	INLI	OPPR	PSDM
No. of (weighted) respondents	9215.3	8779.3	8678.7	7697.9	8622.8	8906.5	8864.4
No. of items	6	1	1	1	1	1	1
Mean measure (Mean score)	0.97 (18.9)	0.93 (3.6)	0.95 (3.5)	1.00 (2.9)	1.02 (3.2)	1.00 (3.7)	0.99 (3.3)
S.D. population (S.D. score)	1.29 (5.7)	1.35 (1.0)	1.33 (1.1)	1.41 (1.1)	1.46 (1.0)	1.31 (1.0)	1.40 (1.1)
Max. measure (Max. score)	5.26 (30.0)	2.83 (5.0)	2.84 (5.0)	3.77 (5.0)	3.82 (5.0)	2.86 (5.0)	3.43 (5.0)
Min. measure (Min. score)	-3.96 (3.0)	-2.66 (1.0)	-2.20 (1.0)	-1.49 (1.0)	-2.08 (1.0)	-2.16 (1.0)	-1.83 (1.0)
Person reliability [Alpha (r_{xx})]	0.77 (0.95)	-	-	-	-	-	-
Person RMSE (SEM)	0.62 (1.27*)	1.37 (-)	1.36 (-)	1.32 (-)	1.35 (-)	1.39 (-)	1.33 (-)
Person separation	1.84	-	-	-	-	-	-
Person strata	2.79	-	-	-	-	-	-
Item mean measure (RMSE)	0.52 (0.01)	0.10 (0.01)	0.29 (0.01)	1.12 (0.01)	0.76 (0.01)	0.19 (0.01)	0.68 (0.01)
Item reliability	1.00	-	-	-	-	-	-
Item Infit MNSQ (Zstd)	0.81 (-9.90)	-	-	-	-	-	-

Item Outfit MNSQ (Zstd)	0.82 (-9.90)	-	-	-	-	-	-

* approximate due to missing data

Measures of Skills Requirements had a person reliability of 0.77, with a person separation of 1.84 and person strata of 2.79. It's corresponding classical test theory reliabilities (i.e., Cronbach's Alpha) was 0.95. Each domain had only one overall measure of Skills Requirements, and as such, no reliabilities and/or separation indices are reported. On average, respondents found the INLI domain (Mean measure = 1.02) easier to endorse and found the COCO domain (Mean measure = 0.93) harder to endorse.

Table 27 Main stage survey psychometric properties of personality traits

Properties	Extraversion	Agreeableness	Conscientious	Emotional stability*	Open-Mindedness
No. of (weighted) respondents	10856.0	10790.9	10868.3	10750.9	10942.8
No. of items	3	3	3	3	3
Mean measure (Mean score)	0.87 (9.0)	0.89 (9.1)	0.95 (11.1)	0.86 (8.4)	0.92 (9.2)
S.D. population (S.D. score)	1.30 (2.4)	1.37 (2.1)	1.46 (2.4)	1.42 (2.6)	1.32 (2.2)
Max. measure (Max. score)	5.30 (15.0)	4.29 (13.0)	4.20 (15.0)	4.90 (14.0)	4.61 (14.0)
Min. measure (Min. score)	-4.15 (2.0)	-5.94 (2.0)	-4.29 (2.0)	-3.18 (2.0)	-4.15 (2.0)
Person reliability [Alpha (r_{xx})]	0.57 (0.73)	0.44 (0.80)	0.52 (0.90)	0.60 (0.80)	0.51 (0.70)
Person RMSE (SEM)	0.85 (1.25*)	1.03 (0.93*)	1.01 (0.79*)	0.89 (1.17*)	0.92 (1.21*)
Person separation	1.15	0.88	1.03	1.23	1.02
Person strata	1.87	1.51	1.71	1.97	1.69

Item mean measure (RMSE)	0.64 (0.01)	-0.41 (0.02)	-0.11 (0.02)	0.72 (0.01)	0.25 (0.01)
Item reliability	1.00	1.00	1.00	1.00	0.99
Item Infit MNSQ (Zstd)	0.72 (-9.90)	0.75 (-9.90)	0.78 (-9.90)	0.60 (-9.90)	0.72 (-9.90)
Item Outfit MNSQ (Zstd)	0.72 (-9.90)	0.75 (-9.90)	0.78 (-9.90)	0.60 (-9.90)	0.72 (-9.90)

* The trait measured was *neuroticism,* but this was reversed to become 'emotional stability' so that all of the personality scales ran in the same direction and could be more easily compared.

Three items were used to measure each personality trait. Their person reliabilities ranged from 0.44 to 0.60, with corresponding person separations ranging from 0.88 to 1.23, and person strata ranging from 1.51 to 1.97. Consequently, each personality trait was sensitive enough to distinguish between 1.5 and 2 levels of endorsement. Their corresponding classical test theory reliability (i.e., Cronbach's Alpha) ranged from 0.70 to 0.90. On average, respondents found the Conscientiousness trait (Mean measure = 0.95) easier to endorse and found the Emotional Stability trait (Mean measure = 0.86) harder to endorse.

4.6 Recalibrating the logit measures for Skills Supply, Skills Requirements and personality traits onto a positive, linear user-friendly scale from 0 to 100

The logit measures were recalibrated into a positive linear scale, ranging from 0 to 100, by multiplying each logit by a predefined user-scaled unit and adding the items' mean difficulty. For example, a logit of 0.91 was transformed into a rescaled measurement of 57.93 [(0.91 × 6.23) + 50.71 = 57.93].

5 Projecting future Skills Supply and Skills Requirements

A core objective of this research was to explore how Skills Supply and Skills Gaps may change between 2023 and 2035 as a result of projected changes in the composition of the population, in the industrial and occupational distribution of employment, and in the skills utilised within each occupation. This is so as to inform solutions and policy responses to *anticipated* future Skills Gaps in the future labour market, which may differ from the Skills Gaps we see today. This section of the Technical Supplement describes how we projected future Skills Supply and Skills Gaps using data from the NFER Essential Employment Skills survey combined with other sources.

5.1 Methodology for projecting *future* Skills Supply and Skills Gaps

The methodology for projecting *future* Skills Supply and Skills Gaps through to 2035 comprised three stages. Stage one involved re-weighting our survey data to account for projected changes in the composition, health, education and working hours of the population, before comparing our 2035 projected data with actual data from 2023. Stage two involved exploring the impact that projected changes in the occupational and industrial distribution of employment are likely to have on Skills Supply and Skills Gaps. The final stage involved anticipating the effects on Skills Gaps of projected changes in the demand for skills within occupations. This enabled us to examine the impact of changes in i) the composition of the population, ii) the jobs that will be available in the future, and iii) the skills that will be needed to do those jobs.

This first required us to project the demographic composition, health status, qualification levels, employment status, and occupational and industrial distribution of employment in the English working-age population in 2035. These projections were then used to re-weight our survey data, enabling us to account for the effects of population changes and changes in the distribution of employment. The characteristics that we forecast – for each of the three subpopulations – were: Age crossed with sex, ITL1 region (former Government Office Regions), Highest qualification (Level 4+ / below Level 4) crossed with age, Economic Activity, Ethnicity, Disability Status (whether is under Equality Act 2010 or not), Country of birth, Occupation (UK SOC 2020 Major Groups) and Industry. These match the list of variables used in the Main stage survey weighting. We forecasted future values of these variables using LFS data from 125 quarters between 1992 Q2 and 2023 Q2 (or from when the variable was first recorded). Where the time series had gaps or large unexplained changes from one year to the next (most likely because of changes in the way variables were defined), we replaced the original data with a simple linear interpolation.

To predict future values of each variable from past values of the same variables, we used Autoregressive integrated moving average (ARIMA) models, in which the value of a time series depends on its own lagged values and the current and past values of a white noise process. Crucially, ARIMA models require time series to be stationary, which means that variances and autocovariances do not depend on the time period at which the series is observed. Hence, a preliminary step involved testing the stationarity of each time series (Stock & Watson, 2019; Verbeek, 2017). To do this, we conducted Augmented Dickey-Fuller tests, progressively adding quarters of lags of the differenced dependent variable to our regression models and identifying the optimal number of lags. If we could not reject the null hypothesis of non-stationary, we differenced the variable and re-ran the diagnostic tests. If we still could not reject the null hypothesis of non-stationary, we took the second differences of the variable and re-ran the diagnostic tests. Partial Autocorrelation (PAC) and Autocorrelation (AC) graphs of the stationary series were produced up to forty lags. These provided guidance in choosing the most appropriate parameters of the ARIMA models, i.e.

the order of the autoregressive component (p) and the order of the moving-average component (q), to forecast each time series 50 quarters ahead (through to 2035 Q4). We tested the quality of competing ARIMA models for prediction by dividing our LFS dataset into a training sample and a test sample, iteratively estimating the Pseudo Out-of-Sample Root Mean Squared Forecast Error (RMSFE). This followed the procedure in Stock & Watson (2019, Sec. 15.5). This enabled us to select the most accurate model for forecasting each variable. Using the optimal model, we made iterated multi-period forecasts up to 2035 Q4 of the undifferenced series. Our resulting forecasts for each variable are displayed in Appendix F. We then re-weighted our survey data to account for projected changes in the population and compared our 2035 projections with actual data from 2023 to examine the effects of projected population changes on Skills Supply and Skills Gaps.

Next, we explored the effects that projected changes in employment are likely to have on skills. Occupational group (SOC) and industry (SIC) were *not* used in our original weighting procedure as people's responses to questions about their occupation and industry are dependent on what prompts they are given. Therefore, we needed to first re-weight our 2023 data adding SOC and SIC to the original weighting variables. We then we re-weighted it again using our 2035 projected population totals together with 2035 employment projections produced earlier for The Skills Imperative 2035 (Wilson *et al.*, 2022). Comparing the two datasets enabled us to explore the potential impact that projected changes in employment might have on EES supply and EES gaps, overall and by occupation. This was treated as *exploratory* given including SOC and SIC in the weighting scheme introduces a new source of error. For more detail on the employment projections utilised in this stage of the process, readers are signposted to *The Skills Imperative 2035: Occupational Outlook – Long run employment prospects for the UK* (Wilson *et al.*, 2022).

Finally, we explored the potential impact on Skills Gaps of projected changes in the skills workers will be required to utilise within each occupation. This stage of the process involved using earlier skills projections produced for The Skills Imperative 2035 (Dickerson *et al.*, 2023). This previous stage of the research programme involved generating projections of the skill requirements for each occupation in 2035. These can be found in *The Skills Imperative 2035: An analysis of the demand for skills in the labour market in 2035* (Dickerson *et al.*, 2023). Where our skills projections indicated that specific skills measured in our survey will be more heavily utilised within an occupation in 2035, we scaled up the Skills Requirement scores of people in that occupational group in our weighted survey sample (without adjusting their Skills Supply, which might – in reality – increase as a response to increased requirements). This enabled us to explore the potential effects of changes in Skills Requirements within occupations on Skills Gaps across the occupational hierarchy.

Our projections should be treated as exploratory and comparisons between Skills Gaps in 2023 and 2035 should be interpreted cautiously.

The analyses that were conducted using our survey data are detailed in a second Technical Supplement, to which readers are signposted.

Bibliography

Autor, D.H., Levy, F. and Murnane, R.J. (2003) 'The skill content of recent technological change: an empirical exploration', *Quarterly Journal of Economics*, 118(4), pp. 1279–1333.

Costa, P. and McCrae, R. (1992) *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa: FL: Psychological Assessment Resources.

Deming, D.J. (2015) *The growing importance of social skills in the labor market*. 21473. Available at: https://www.nber.org/system/files/working_papers/w21473/w21473.pdf (Accessed: 19 February 2024).

Dickerson, A., Rossi, G., Bocock, L., Hillary, J. and Simcock, D. (2023) *The Skills Imperative 2035: An analysis of the demand for skills in the labour market in 2035*. Available at: https://www.nfer.ac.uk/media/utwf2grb/the_skills_imperative_2035_working_paper_3_an_analysis_of_the_demand_for_skills_in_the_labour_market_in_2035.pdf (Accessed: 19 February 2024).

European Commission (2019) *Key competences for lifelong learning*. Available at: https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/%20language-en (Accessed: 19 February 2024).

European Commission, Directorate General for Employment, Social Affairs and Inclusion, and the US Department of Labor/Employment and Training Administration (2022) *The crosswalk between ESCO and O*NET. Technical Report*. Available at: https://esco.ec.europa.eu/system/files/2022-12/ONET%20ESCO%20Technical%20Report.pdf (Accessed: 13 March 2024).

Felstead, A., Gallie, D., Green, F. and Zhou, Y. (2007) *Skills at work, 1986 to 2006*. Oxford: ESRC Centre on Skills, Knowledge and Organisational Performance.

Grol-Prokopczyk, H. (2018) 'In pursuit of anchoring vignettes that work: evaluating generality versus specificity in vignette texts', *The Journals of Gerontology: Series B*, 73(1), pp. 54–63. Available at: https://doi.org/10.1093/geronb/gbx048.

He, J., Van de Vijver, F.J.R., Fetvadjiev, V.H., de Carmen Dominguez Espinosa, A., Adams, B., Alonso–Arbiol, I., Aydinli–Karakulak, A., Buzea, C., Dimitrova, R., Fortin, A., Hapunda, G., Ma, S., Sargautyte, R., Sim, S., Schachner, M.K., Suryani, A., Zeinoun, P. and Zhang, R. (2017) 'On enhancing the cross–cultural comparability of Likert–Scale Personality and Value Measures: a comparison of common procedures', *European Journal of Personality*, 31(6), pp. 642–657. Available at: https://doi.org/10.1002/per.2132.

Hoskins, B. and Liu, D. L. (2019) *Professor Bryony Hoskins University of Roehampton, London, U.K.* Available at: https://www.unicef.org/mena/media/7011/file/Measuring%20life%20skills_web.pdf.pdf (Accessed: 12 March 2024).

Kline, P. (1993) *Personality: The Psychometric View*. London: Routledge.

Linacre, J. (2004) 'Optimizing rating scale category effectiveness', in E.V. Jr. Smith and R. M. Smith (eds) *Introduction to Rasch measurement theory: Models and applications*. London: JAM Press.

Linacre, J.M. (2023a) *A User's Guide to Winsteps® Rasch-Model computer programs*. Available at: https://www.winsteps.com/winman/copyright.htm (Accessed: 13 March 2024).

Linacre, J.M. (2023b) *Winsteps® (Version 5.4.1.0) [Computer Software]*. Available at: https://www.winsteps.com/index.htm (Accessed: 13 March 2024).

Lira, B., O'Brien, J.M., Peña, P.A., Galla, B.M., D'Mello, S., Yeager, D.S., Defnet, A., Kautz, T., Munkacsy, K. and Duckworth, A.L. (2022) 'Large studies reveal how reference bias limits policy applications of self-report measures', *Scientific Reports*, 12(1), p. 19189. Available at: https://doi.org/10.1038/s41598-022-23373-9.

Lucas, R.E. (2018) 'Reevaluating the strengths and weaknesses of self-report measures of subjective well-being', in E. Diener, S. Oishi, and L. Tay (eds) *Handbook of well-being*. Salt Lake City, UT: DEF Publishers., pp. 45–55. Available at: https://florianesilvestri.files.wordpress.com/2019/07/handbook-of-well-being.pdf (Accessed: 20 February 2024).

Marfeo, E., Ni, P., Chan, L., Rasch, E.K. and Jette, A.M. (2014) 'Combining agreement and frequency rating scales to optimize psychometrics in measuring behavioral health functioning', *Journal of clinical epidemiology*, 67(7), pp. 781–784. Available at: https://doi.org/10.1016/j.jclinepi.2014.02.005.

Organisation for Economic Co-operation and Development (2013) *Technical report of the Survey of Adult Skills (PIAAC)*. Available at: https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf (Accessed: 20 February 2024).

Office for National Statistics (2018) 'Personal well-being in the UK QMI'. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingintheukqmi (Accessed: 20 February 2024).

Paulhus, D.L. and Vazire, S. (2007) 'The self-report method', in R.W. Robins, R.C. Fraley, and R.F. Krueger (eds) *The handbook of research methods in personality*, pp. 224–239. Available at: https://www2.psych.ubc.ca/~dpaulhus/research/SDR/downloads/CHAPTERS/2008%20Handbook%20Research%20Methods/paulhus-vazire%2007%20chap.pdf (Accessed: 20 February 2024).

Popov, D., Snelson, S. and Baily, T. (2022) *Review of skills taxonomies*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1078099/Review_of_skills_taxonomies_report_prepared_for_the_SPB_May_2022.pdf (Accessed: 20 February 2024).

Rasch, G. (1960) *Probabilistic models in some intelligence and attainment test.* (Expanded edition, Chicago: The University of Chicago Press, 1980). Copenhagen: Danish Institute for Educational Research.

Ravenscroft, T. and Baker, L. (2020) *Towards a Universal Framework for Essential Skills. A Review of the Skills Builder Framework. Final Report for the Essential Skills Taskforce.* Available at: https://global-uploads.webflow.com/5a86b2cd68b41700017162ca/5ec2b44c1cac8f93cea68444_Essential%20Skills%20Taskforce%20Report%20-%20Final%20(May%202020).pdf (Accessed: 13 March 2024).

Sayer, A. (2007) ‘Dignity at work: broadening the agenda’, *Organization*, 14(4), pp. 565–581. Available at: https://doi.org/10.1177/1350508407078053.

Soto, C.J. and John, O.P. (2017) ‘The next Big Five Inventory (BFI-2): developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power.’, *Journal of Personality and Social Psychology*, 113(1), pp. 117–143. Available at: https://doi.org/10.1037/pspp0000096.

Steponavičius, M., Gress-Wright, C. and Linzarini, A. (2023) *Social and emotional skills (SES): Latest evidence on teachability and impact on life outcomes*. OECD Education Working Papers 304. Available at: https://doi.org/10.1787/ba34f086-en.

Stock, J. and Watson, M. (2019) *Introduction to Econometrics, Global Edition*. 4th edn. London: Pearson.

Taylor, A., Nelson, J., O’Donnell, S., Davies, E. and Hillary, J. (2022) *The Skills Imperative 2035: what does the literature tell us about essential skills most needed for work?* Available at: https://www.nfer.ac.uk/media/hgwls32i/the_skills_imperative_2035_working_paper_1_essential_skills_literature_review.pdf (Accessed: 19 February 2024).

Tong, J., Bickmeier, R.M. and Rogelberg, S.G. (2020) ‘A comparison of frequency- and agreement-based response formats in the measurement of burnout and engagement’, *International Journal of Environmental Research and Public Health*, 17(2), p. 543. Available at: https://doi.org/10.3390/ijerph17020543.

United Nations Children's Fund Middle East and North Africa (UNICEF MENA) (2017) *Reimagining Life Skills and Citizenship Education in the Middle East and North Africa. A Four-Dimensional and Systems Approach to 21st Century Skills.* Available at: https://www.unicef.org/mena/media/6151/file/LSCE%20Conceptual%20and%20Programmatic%20Framework_EN.pdf%20.pdf (Accessed: 9 May 2023).

Verbeek, M. (2017) *A guide to modern econometrics*. 5th edn. Hoboken, NJ: John Wiley & Sons Inc.

Wilson, R., Bosworth, D., Bosworth, L., Cardenas-Rubio, J., Day, R., Patel, S., Bui, H., Seymour, D., Thoung, C. and Hillary, J. (2022) *The Skills Imperative 2035: Occupational Outlook – Long-run employment prospects for the UK*. Available at: https://www.nfer.ac.uk/media/ybkcqvuz/the_skills_imperative_2035_working_paper_2_headline_report.pdf (Accessed: 19 February 2024).

Wright, B.D. and Stone, M. (1979) *Best Test Design: Rasch Measurement*. Chicago, IL: MESA Press. Available at: https://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=measurement (Accessed: 13 March 2024).

Appendix A: Weighting targets and weighted sample profile for the combined data

Variable / Sub-population	Category	%
Sex by age		
Paid adult workers and young people working more than 20h/week	Male 15-18	0.2%
	Male 19-29	9.3%
	Male 30-39	10.4%
	Male 40-49	9.2%
	Male 50-65	14.0%
	Female 15-18	0.2%
	Female 19-29	8.6%
	Female 30-39	9.7%
	Female 40-49	8.8%
	Female 50-65	13.2%
Young people yet to enter the full labour market	Male 15-18	1.9%
	Female 15-18	1.9%
Adults never employed and long term unemployed	Male 19-29	1.6%
	Male 30-39	0.5%
	Male 40-49	0.5%
	Male 50-65	1.9%
	Female 19-29	1.8%
	Female 30-39	1.2%
	Female 40-49	1.2%
	Female 50-65	3.3%

Not able to categorise	Male 15-29	0.2%
	Male 30-49	0.1%
	Male 50-65	0.1%
	Female 15-29	0.1%
	Female 30-49	0.1%
	Female 50-65	0.1%
Region		
Paid adult workers and young people working more than 20h/week	North East/North West/Yorkshire & Humber	22.2%
	East Midlands/West Midlands/East of England	25.0%
	London	14.6%
	South East & South West	21.9%
Young people yet to enter the full labour market	North East/North West/Yorkshire & Humber	1.0%
	East Midlands/West Midlands/East of England	1.1%
	London	0.7%
	South East & South West	1.0%
Adults never employed and long term unemployed	North East/North West/Yorkshire & Humber	3.8%
	East Midlands/West Midlands/East of England	3.4%
	London	2.1%
	South East & South West	2.5%
Not able to categorise	North East/North West/Yorkshire & Humber	0.1%
	East Midlands/West Midlands/East of England	0.2%

	London	0.2%
	South East & South West	0.1%
Highest qualification		
Paid adult workers and young people working more than 20h/week	No level 4+	41.0%
	Level 4+	42.7%
Young people yet to enter the full labour market	All	3.7%
Adults never employed and long term unemployed	No level 4+	9.0%
	Level 4+	2.9%
Not able to categorise	All	0.7%
Ethnicity		
Paid adult workers and young people working more than 20h/week	White: British	61.9%
	White: Other	7.5%
	Asian: Indian	3.2%
	Asian: Pakistani	1.6%
	Asian: Other	2.7%
	Black	3.4%
	Other/mixed	3.5%
Young people yet to enter the full labour market	White: British	2.6%
	White: Other	0.2%
	Asian: Indian	0.1%
	Asian: Pakistani	0.2%
	Asian: Other	0.1%
	Black	0.3%
	Other/mixed	0.3%

Adults never employed and long term unemployed	White: British	8.3%
	White: Other	0.6%
	Asian: Indian	0.4%
	Asian: Pakistani	0.6%
	Asian: Other	0.7%
	Black	0.7%
	Other/mixed	0.6%
Not able to categorise	All	0.7%
Disability / long-term illness (Equalities Act definition)		
Paid adult workers and young people working more than 20h/week	No	68.4%
	EA disability	15.4%
Young people yet to enter the full labour market	No & all 15 y olds	3.2%
	EA disability	0.5%
Adults never employed and long term unemployed	No	5.6%
	EA disability	6.3%
Not able to categorise	All	0.7%
Economic Activity		
Paid adult workers and young people working more than 20h/week	FT employee	48.5%
	PT employee (20-34 hours)	11.5%
	PT employee (<20 hours)	5.3%
	Self-employed	9.7%
	Other	8.8%
Young people yet to enter the full labour market	All	3.7%
	Unemployed	0.9%

Adults never employed and long term unemployed	Full time student	1.9%
	Retired	1.3%
	Other economically inactive	7.8%
Not able to categorise	All	0.7%
Country of birth		
Paid adult workers and young people working more than 20h/week	UK	65.5%
	Not UK	18.2%
Young people yet to enter the full labour market	UK	3.3%
	Not UK	0.4%
Adults never employed and long term unemployed	UK	9.1%
	Not UK	2.8%
Not able to categorise	All	0.7%

Appendix B Main study Item indices

Entry number	Total Score	Total Count	Measure (S.E.)	Infit MNSQ (Zstd)	Outfit MNSQ (Zstd)	Item
1	41678.68	11063.54	-0.22 (0.01)	0.83 (-9.90)	0.80 (-9.90)	PSDM1_01
2	41348.43	11037.88	-0.24 (0.01)	0.86 (-9.90)	0.84 (-9.90)	PSDM2_02
3	34632.89	10069.51	0.02 (0.01)	1.02 (1.38)	0.99 (-0.51)	PSDM3_03
4	DELETED	DELETED	DELETED	DELETED	DELETED	PSDM4_04
5	42661.7	11112.86	-0.48 (0.01)	0.81 (-9.90)	0.79 (-9.90)	PSDM5_05
6	26565.17	8527.497	1.00 (0.02)	1.02 (1.56)	1.02 (1.23)	PSDM6_R_06
7	41448.85	10999.31	-0.29 (0.01)	0.86 (-9.90)	0.83 (-9.90)	PSDM7_07
8	33460.95	9234.615	-0.20 (0.01)	1.02 (1.09)	1.02 (1.02)	PSDM8_R_08
9	40675.94	11105.96	-0.21 (0.01)	0.81 (-9.90)	0.79 (-9.90)	PSDM9_09
10	42535.98	11085.85	-0.37 (0.01)	0.80 (-9.90)	0.77 (-9.90)	PSDM10_10
11	38694.88	10967.57	0.14 (0.01)	0.89 (-9.31)	0.86 (-9.90)	PSDM11_11
12	40196.33	11079.54	-0.03 (0.01)	0.80 (-9.90)	0.78 (-9.90)	PSDM12_12
13	42570.33	11056.92	-0.24 (0.01)	0.79 (-9.90)	0.75 (-9.90)	OPPR_1_13
14	42192.67	10986.7	-0.37 (0.01)	0.84 (-9.90)	0.82 (-9.90)	OPPR_2_14
15	39235.84	10924.22	0.11 (0.01)	0.90 (-8.15)	0.87 (-9.05)	OPPR_3_15
16	38163.53	10841.56	0.11 (0.01)	0.95 (-4.12)	0.92 (-5.47)	OPPR_4_16
17	33454.45	10769.01	0.69 (0.01)	0.98 (-1.39)	0.96 (-2.69)	OPPR_5_17
18	43533.15	10970.96	-0.37 (0.01)	0.83 (-9.90)	0.78 (-9.90)	OPPR_6_18

19	33680.34	10239.23	0.37 (0.01)	1.17 (9.90)	1.14 (9.90)	OPPR_7_19
20	43290.29	10791.72	-0.52 (0.01)	0.92 (-6.33)	0.87 (-8.47)	OPPR_8_20
21	37776.53	10646.78	0.31 (0.01)	0.98 (-1.36)	0.93 (-4.17)	OPPR_9_21
22	23679.04	8421.228	1.29 (0.01)	1.11 (7.32)	1.11 (7.13)	OPPR_10_R_22
23	37185.45	10773.47	0.28 (0.01)	1.01 (0.84)	0.98 (-1.60)	OPPR_11_23
24	37074.8	10752.94	0.29 (0.01)	0.97 (-2.13)	0.95 (-3.86)	OPPR_12_24
25	33321.62	10809.89	-0.51 (0.02)	0.92 (-5.83)	0.92 (-6.14)	COMM_1_25
26	28039.11	10088.49	0.36 (0.01)	1.09 (6.21)	1.08 (5.82)	COMM_2_26
27	31435.05	10528.61	-0.10 (0.02)	0.97 (-2.06)	0.96 (-2.94)	COMM_3_27
28	31585.07	10082.17	-0.61 (0.02)	1.10 (7.37)	1.09 (6.56)	COMM_4_28
29	32007.48	10833.08	-0.23 (0.02)	0.96 (-3.14)	0.95 (-3.76)	COMM_5_29
30	31904	10475.49	-0.24 (0.02)	0.99 (-0.53)	0.97 (-1.79)	COMM_6_30
31	27974.14	10546.61	0.67 (0.01)	1.03 (2.61)	1.02 (1.54)	COMM_7_31
32	33059.54	10683.33	-0.16 (0.01)	0.96 (-2.70)	0.93 (-4.74)	COMM_8_32
33	33103.07	10869.5	-0.34 (0.02)	0.91 (-6.81)	0.90 (-7.68)	COMM_9_33
34	31521.07	10360.33	-0.28 (0.02)	0.96 (-2.97)	0.94 (-3.97)	COMM_10_34
35	32466.85	10848.93	-0.32 (0.02)	0.96 (-2.76)	0.96 (-3.04)	COMM_11_35
36	DELETED	DELETED	DELETED	DELETED	DELETED	COMM_12_36
37	DELETED	DELETED	DELETED	DELETED	DELETED	COCO_1_37
38	36214.93	10847.73	-1.51 (0.02)	0.95 (-3.83)	0.94 (-4.57)	COCO_2_38
39	28749	10261.77	1.29 (0.01)	1.16 (9.90)	1.15 (9.90)	COCO_3_R_39
40	30034.25	10404.75	0.06 (0.02)	1.00 (-0.17)	0.98 (-1.14)	COCO_4_40

41	34726.51	10869.36	-0.94 (0.02)	0.97 (-2.43)	0.96 (-2.81)	COCO_5_41
42	35311.58	11021.2	-1.32 (0.02)	0.92 (-6.20)	0.92 (-6.26)	COCO_6_42
43	33934.65	10774.79	-0.78 (0.02)	0.93 (-5.43)	0.92 (-6.16)	COCO_7_43
44	DELETED	DELETED	DELETED	DELETED	DELETED	COCO_8_44
45	DELETED	DELETED	DELETED	DELETED	DELETED	COCO_9_45
46	41360.61	10498.52	-0.57 (0.01)	1.03 (2.38)	1.01 (0.55)	COCO_10_46
47	24796.37	9867.527	1.54 (0.01)	1.24 (9.90)	1.22 (9.90)	COCO_11_R_47
48	DELETED	DELETED	DELETED	DELETED	DELETED	COCO_12_48
49	35317.04	11032.7	0.43 (0.01)	0.95 (-4.28)	0.93 (-5.14)	TCRE_1_49
50	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_2_50
51	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_3_51
52	32988.88	10821.25	0.72 (0.01)	1.14 (9.90)	1.12 (8.25)	TCRE_4_52
53	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_5_53
54	35156.04	10898.74	0.51 (0.01)	1.07 (5.90)	1.05 (3.59)	TCRE_6_54
55	37640.98	11109.66	0.14 (0.01)	0.94 (-5.18)	0.92 (-5.89)	TCRE_7_55
56	32527.43	10862.87	0.66 (0.01)	1.10 (7.78)	1.09 (6.49)	TCRE_8_56
57	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_9_57
58	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_10_58
59	DELETED	DELETED	DELETED	DELETED	DELETED	TCRE_11_59
60	22271.23	7826.77	-0.28 (0.02)	1.03 (1.53)	1.04 (2.12)	TCRE_12_R_60
61	32778.86	10532.04	-0.46 (0.02)	0.97 (-2.68)	0.95 (-3.60)	INLI_1_61
62	DELETED	DELETED	DELETED	DELETED	DELETED	INLI_2_62

63	34628.65	10907.77	-0.85 (0.02)	0.87 (-9.72)	0.86 (-9.90)	INLI_3_63
64	33559.24	10788.68	-0.63 (0.02)	0.95 (-3.84)	0.94 (-4.81)	INLI_4_64
65	DELETED	DELETED	DELETED	DELETED	DELETED	INLI_5_65
66	33770.38	10835.86	-0.93 (0.02)	0.91 (-7.11)	0.90 (-7.71)	INLI_6_66
67	DELETED	DELETED	DELETED	DELETED	DELETED	INLI_7_67
68	31712.39	10001.16	-0.80 (0.02)	1.03 (2.00)	1.01 (0.58)	INLI_8_68
69	35201.37	10950.93	-1.33 (0.02)	0.91 (-7.41)	0.90 (-8.06)	INLI_9_69
70	31545.66	10572.86	-0.27 (0.02)	1.00 (-0.23)	0.99 (-0.79)	INLI_10_70
71	30182.18	10659.42	0.17 (0.01)	1.03 (2.63)	1.03 (2.33)	INLI_11_71
72	33653.73	10840.51	-0.54 (0.02)	0.89 (-8.96)	0.87 (-9.65)	INLI_12_72
73	28957.79	8864.352	0.68 (0.01)	1.00 (-0.06)	1.01 (0.87)	O*NET_PSDM_73
74	32782.92	8906.531	0.19 (0.01)	0.90 (-6.94)	0.90 (-6.57)	O*NET_OPPR_74
75	30771.44	8678.654	0.29 (0.01)	0.96 (-2.68)	0.97 (-2.20)	O*NET_COMM_75
76	31731.45	8779.345	0.10 (0.01)	0.98 (-1.23)	0.98 (-1.08)	O*NET_COCO_76
77	22345.58	7697.894	1.12 (0.01)	1.09 (5.96)	1.09 (5.89)	O*NET_TCRE_77
78	27670.88	8622.753	0.76 (0.01)	0.98 (-1.61)	1.00 (-0.29)	O*NET_INLI_78
79	29431.82	10249.07	1.05 (0.01)	1.27 (9.90)	1.28 (9.90)	BIG_5_11_R_79
80	33211.89	10583.19	0.73 (0.01)	1.17 (9.90)	1.17 (9.90)	BIG_5_21_80
81	35576.93	10471.24	0.14 (0.01)	1.08 (5.94)	1.09 (6.44)	BIG_5_31_81
82	32367.95	10223.26	-1.36 (0.02)	1.15 (9.90)	1.15 (9.90)	BIG_5_12_82
83	36932.66	10197.23	-0.24 (0.01)	1.21 (9.90)	1.20 (9.90)	BIG_5_22_R_83
84	28482.35	10200.39	0.35 (0.02)	1.10 (6.51)	1.09 (5.89)	BIG_5_32_84

85	38569.41	10300.88	-0.17 (0.01)	1.12 (8.18)	1.09 (6.25)	BIG_5_13_R_85
86	36352.17	10511.72	0.21 (0.01)	1.13 (9.90)	1.13 (9.35)	BIG_5_23_R_86
87	46092.57	10546.17	-0.38 (0.02)	0.98 (-1.76)	0.98 (-1.59)	BIG_5_33_87
88	26034.56	10156.81	1.43 (0.01)	1.32 (9.90)	1.31 (9.90)	BIG_5_14_R_88
89	36100.4	10343.11	0.19 (0.01)	1.16 (9.90)	1.16 (9.90)	BIG_5_24_R_89
90	28350.37	10347.17	0.53 (0.01)	1.09 (6.68)	1.07 (5.27)	BIG_5_34_90
91	36458.06	10347.24	0.14 (0.01)	1.32 (9.90)	1.31 (9.90)	BIG_5_15_91
92	35648.02	10406.7	0.10 (0.01)	1.12 (9.30)	1.13 (9.52)	BIG_5_25_R_92
93	28972.02	10693.8	0.51 (0.01)	0.99 (-0.38)	0.99 (-0.56)	BIG_5_35_93

Appendix C Main stage study item descriptions

Entry number	Item	Description of item
1	PSDM1_01	I think about a situation in detail before making a decision.
2	PSDM2_02	I look for extra information to help me make a decision.
3	PSDM3_03	When I have an important decision to make, I ask others for guidance.
4	PSDM4_04	I think about different potential solutions to a problem.
5	PSDM5_05	I think about the different options before making a decision.
6	PSDM6_R_06	My emotions tend to dominate when I make decisions.
7	PSDM7_07	When I am thinking about a solution to a problem, I consider the possible risks.
8	PSDM8_R_08	When approaching a problem, I do the first thing that comes into my head.
9	PSDM9_09	When I am tackling a problem, I look at different possible outcomes.
10	PSDM10_10	When a proposed solution to a problem is not working, I try to understand why.
11	PSDM11_11	Once I have solved a problem, I think about how well the solution has worked.
12	PSDM12_12	I approach problems by thinking about the benefits of possible solutions.
13	OPPR_1_13	I consider the steps needed to achieve a goal.
14	OPPR_2_14	When things change, I can adjust my plans.
15	OPPR_3_15	I break tasks down into steps to help me check my progress.
16	OPPR_4_16	I find feedback from others helps me work towards my goal.
17	OPPR_5_17	I make a timeline to help achieve a goal.
18	OPPR_6_18	I can identify the most important tasks for the day ahead.
19	OPPR_7_19	I finish a task before moving on to the next one.
20	OPPR_8_20	I prioritise important tasks over less important ones.
21	OPPR_9_21	I make notes to keep track of my tasks.
22	OPPR_10_R_22	At busy times, I tackle whatever task comes to mind first.
23	OPPR_11_23	I enjoy making plans.
24	OPPR_12_24	When I have a lot going on, I order my thoughts by importance.
25	COMM_1_25	When communicating with someone, I try to see things from their point of view.
26	COMM_2_26	I use my body language to reinforce what I am trying to say.
27	COMM_3_27	When speaking with others, I know when it is my turn to talk.

28	COMM_4_28	I change the way I talk to someone based on my relationship with them.
29	COMM_5_29	I make sure I understand what a person means before responding to them.
30	COMM_6_30	I think about how to express my thoughts before I write something down.
31	COMM_7_31	I enjoy communicating with someone whose views are different from mine.
32	COMM_8_32	I am confident asking questions when something doesn't make sense to me.
33	COMM_9_33	In discussion, I use facts to support my beliefs.
34	COMM_10_34	Depending on the situation, I consider whether to speak or write to someone.
35	COMM_11_35	When someone asks me for information, I know where to look for it.
36	COMM_12_36	I adapt the way I communicate information depending on who it is for.
37	COCO_1_37	I make an effort to respect the opinions of people around me.
38	COCO_2_38	I take into account that each person has different needs.
39	COCO_3_R_39	I dislike it when people challenge my views.
40	COCO_4_40	If other people disagree with me, I try to find out why.
41	COCO_5_41	Learning from other people helps me get things done.
42	COCO_6_42	I consider others' ideas and thoughts.
43	COCO_7_43	I discuss my ideas with other people.
44	COCO_8_44	When working with others, people say that I am reliable.
45	COCO_9_45	I listen to other people's points of view.
46	COCO_10_46	I treat everyone the same whether I like them or not.
47	COCO_11_R_47	I do things without consulting other people.
48	COCO_12_48	I am considerate of others when doing something together.
49	TCRE_1_49	I try different approaches in order to solve challenges.
50	TCRE_2_50	I like to invent new things.
51	TCRE_3_51	I use my imagination to help me come up with new ideas.
52	TCRE_4_52	I imagine myself in different situations.
53	TCRE_5_53	I come up with different ideas to improve something.
54	TCRE_6_54	I am curious about unusual ideas.
55	TCRE_7_55	When I am stuck, I find alternative ways to do things.
56	TCRE_8_56	I try my own ideas rather than copying others.
57	TCRE_9_57	I come up with new ideas.
58	TCRE_10_58	I explore different ideas.
59	TCRE_11_59	I create something new by combining different ideas.

60	TCRE_12_R_60	I rely on others to suggest alternative ideas.
61	INLI_1_61	I make judgements based on evidence rather than opinions.
62	INLI_2_62	When making a choice, I weigh up available information.
63	INLI_3_63	I can weigh up the pros and cons of other people's suggestions.
64	INLI_4_64	I can identify when something is presented in a one-sided way.
65	INLI_5_65	It is important to me that I can back up my views with information.
66	INLI_6_66	I know how to check whether information is reliable.
67	INLI_7_67	I can separate facts from opinions.
68	INLI_8_68	There are no right or wrong answers to some questions.
69	INLI_9_69	Before I take action, I make sure the information I use is correct.
70	INLI_10_70	I question other people's views if there is little evidence to support them.
71	INLI_11_71	I can spot flaws in other people's thinking.
72	INLI_12_72	I compare other sources when information seems incorrect.
73	O*NET_PSDM_73	How important is MAKING DECISIONS AND SOLVING PROBLEMS to the performance of your current job? What level of MAKING DECISIONS AND SOLVING PROBLEMS is needed to perform *your current job*?
74	O*NET_OPPR_74	How important is ORGANIZING, PLANNING, AND PRIORITIZING WORK to the performance of your current job? What level of ORGANIZING, PLANNING, AND PRIORITIZING WORK is needed to perform your current job?
75	O*NET_COMM_75	How important is COMMUNICATING WITH SUPERVISORS, PEERS, OR SUBORDINATES to the performance of your current job? What level of COMMUNICATING WITH SUPERVISORS, PEERS, OR SUBORDINATES is needed to perform your current job?
76	O*NET_COCO_76	How important is ESTABLISHING AND MAINTAINING INTERPERSONAL RELATIONSHIPS to the performance of your current job? What level of ESTABLISHING AND MAINTAINING INTERPERSONAL RELATIONSHIPS is needed to perform *your current* job?
77	O*NET_TCRE_77	How important is THINKING CREATIVELY to the performance of your current job? What level of THINKING CREATIVELY is needed to perform your current job?
78	O*NET_INLI_78	How important is CRITICAL THINKING to the performance of your current job?

		What level of CRITICAL THINKING is needed to perform your current job?
79	BIG_5_11_R_79	I tend to be quiet.
80	BIG_5_21_80	I am dominant, act as a leader.
81	BIG_5_31_81	I am full of energy.
82	BIG_5_12_82	I am compassionate, have a soft heart.
83	BIG_5_22_R_83	I am sometimes rude to others.
84	BIG_5_32_84	I assume the best about people.
85	BIG_5_13_R_85	I tend to be disorganized.
86	BIG_5_23_R_86	I have difficulty getting started on tasks.
87	BIG_5_33_87	I am reliable, can always be counted on.
88	BIG_5_14_R_88	I worry a lot.
89	BIG_5_24_R_89	I tend to feel depressed, blue.
90	BIG_5_34_90	I am emotionally stable, not easily upset.
91	BIG_5_15_91	I am fascinated by art, music, or literature.
92	BIG_5_25_R_92	I have little interest in abstract ideas.
93	BIG_5_35_93	I am original, come up with new ideas.

Appendix D Main stage study item categories

Entry number	Item	Categories				
1	PSDM1_01	Never/Almost Never	Sometimes	Often	Almost Always	Always
2	PSDM2_02	Never/Almost Never	Sometimes	Often	Almost Always	Always
3	PSDM3_03	Never/Almost Never	Sometimes	Often	Almost Always	Always
4	PSDM4_04	DELETED				
5	PSDM5_05	Never/Almost Never	Sometimes	Often	Almost Always	Always
6	PSDM6_R_06	Never/Almost Never	Sometimes	Often	Almost Always	Always
7	PSDM7_07	Never/Almost Never	Sometimes	Often	Almost Always	Always
8	PSDM8_R_08	Never/Almost Never	Sometimes	Often	Almost Always	Always
9	PSDM9_09	Never/Almost Never	Sometimes	Often	Almost Always	Always
10	PSDM10_10	Never/Almost Never	Sometimes	Often	Almost Always	Always
11	PSDM11_11	Never/Almost Never	Sometimes	Often	Almost Always	Always
12	PSDM12_12	Never/Almost Never	Sometimes	Often	Almost Always	Always
13	OPPR_1_13	Never/Almost Never	Sometimes	Often	Almost Always	Always
14	OPPR_2_14	Never/Almost Never	Sometimes	Often	Almost Always	Always
15	OPPR_3_15	Never/Almost Never	Sometimes	Often	Almost Always	Always
16	OPPR_4_16	Never/Almost Never	Sometimes	Often	Almost Always	Always
17	OPPR_5_17	Never/Almost Never	Sometimes	Often	Almost Always	Always
18	OPPR_6_18	Never/Almost Never	Sometimes	Often	Almost Always	Always
19	OPPR_7_19	Never/Almost Never	Sometimes	Often	Almost Always	Always
20	OPPR_8_20	Never/Almost Never	Sometimes	Often	Almost Always	Always
21	OPPR_9_21	Never/Almost Never	Sometimes	Often	Almost Always	Always

22	OPPR_10_R_22	Never/Almost Never	Sometimes	Often	Almost Always	Always
23	OPPR_11_23	Never/Almost Never	Sometimes	Often	Almost Always	Always
24	OPPR_12_24	Never/Almost Never	Sometimes	Often	Almost Always	Always
25	COMM_1_25	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
26	COMM_2_26	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
27	COMM_3_27	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
28	COMM_4_28	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
29	COMM_5_29	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
30	COMM_6_30	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
31	COMM_7_31	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
32	COMM_8_32	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
33	COMM_9_33	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
34	COMM_10_34	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
35	COMM_11_35	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
36	COMM_12_36	DELETED				
37	COCO_1_37	DELETED				
38	COCO_2_38	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
39	COCO_3_R_39	Strongly Disagree/Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree

40	COCO_4_40	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
41	COCO_5_41	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
42	COCO_6_42	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
43	COCO_7_43	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
44	COCO_8_44	DELETED				
45	COCO_9_45	DELETED				
46	COCO_10_46	Strongly Disagree/Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
47	COCO_11_R_47	Strongly Disagree/Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
48	COCO_12_48	DELETED				
49	TCRE_1_49	Never/Almost Never	Sometimes	Often	Almost Always	Always
50	TCRE_2_50	DELETED				
51	TCRE_3_51	DELETED				
52	TCRE_4_52	Never/Almost Never	Sometimes	Often	Almost Always	Always
53	TCRE_5_53	DELETED				
54	TCRE_6_54	Never/Almost Never	Sometimes	Often	Almost Always	Always
55	TCRE_7_55	Never/Almost Never	Sometimes	Often	Almost Always	Always
56	TCRE_8_56	Never/Almost Never	Sometimes	Often	Almost Always	Always
57	TCRE_9_57	DELETED				
58	TCRE_10_58	DELETED				
59	TCRE_11_59	DELETED				
60	TCRE_12_R_60	Never/Almost Never		Sometimes	Often	Almost Always / Always
61	INLI_1_61	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree

62	INLI_2_62	DELETED				
63	INLI_3_63	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
64	INLI_4_64	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
65	INLI_5_65	DELETED				
66	INLI_6_66	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
67	INLI_7_67	DELETED				
68	INLI_8_68	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
69	INLI_9_69	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
70	INLI_10_70	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
71	INLI_11_71	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
72	INLI_12_72	Strongly Disagree/Disagree/Slightly Disagree		Slightly Agree	Agree	Strongly Agree
73	O*NET_PSDM_73	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
74	O*NET_OPPR_74	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
75	O*NET_COMM_75	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
76	O*NET_COCO_76	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
77	O*NET_TCRE_77	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
78	O*NET_INLI_78	Score 3 – 4	Score 5 – 6	Score 7 – 8	Score 9 – 10	Score 11 – 12
79	BIG_5_11_R_79	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

80	BIG_5_21_80	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
81	BIG_5_31_81	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
82	BIG_5_12_82	Strongly disagree Disagree		Neither agree nor disagree	Agree	Strongly agree
83	BIG_5_22_R_83	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
84	BIG_5_32_84	Strongly disagree Disagree		Neither agree nor disagree	Agree	Strongly agree
85	BIG_5_13_R_85	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
86	BIG_5_23_R_86	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
87	BIG_5_33_87	Neither agree nor disagree		Agree	Strongly agree	
88	BIG_5_14_R_88	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
89	BIG_5_24_R_89	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
90	BIG_5_34_90	Strongly disagree Disagree		Neither agree nor disagree	Agree	Strongly agree
91	BIG_5_15_91	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
92	BIG_5_25_R_92	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
93	BIG_5_35_93	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

Appendix E Main stage study item thresholds

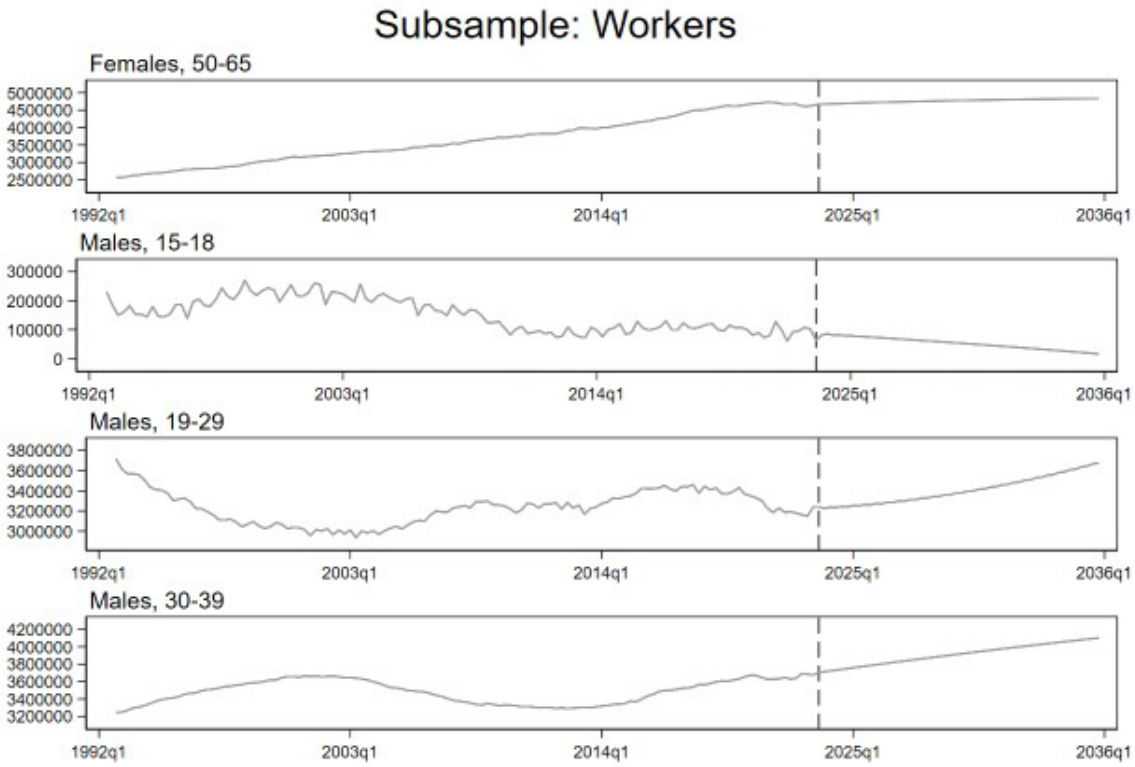
Entry number	Item	1	2	3	4	5
1	PSDM1_01	0	-1.95	-0.06	0.7	1.32
2	PSDM2_02	0	-2.11	-0.03	0.78	1.36
3	PSDM3_03	0	-2.46	0.03	0.95	1.47
4	PSDM4_04	DELETED				
5	PSDM5_05	0	-2.35	-0.2	1.03	1.52
6	PSDM6_R_06	0	-2.72	-1.56	0.42	3.87
7	PSDM7_07	0	-2.08	-0.16	0.92	1.32
8	PSDM8_R_08	0	-2.48	-0.97	0.57	2.89
9	PSDM9_09	0	-2.34	-0.18	0.91	1.61
10	PSDM10_10	0	-2.01	-0.27	0.96	1.32
11	PSDM11_11	0	-1.69	-0.15	0.64	1.21
12	PSDM12_12	0	-1.86	-0.2	0.72	1.35
13	OPPR_1_13	0	-1.59	-0.2	0.85	0.94
14	OPPR_2_14	0	-2.07	-0.16	0.81	1.43
15	OPPR_3_15	0	-1.49	-0.11	0.63	0.97
16	OPPR_4_16	0	-1.79	-0.03	0.79	1.02
17	OPPR_5_17	0	-0.86	-0.15	0.39	0.62
18	OPPR_6_18	0	-1.51	-0.27	0.65	1.13
19	OPPR_7_19	0	-1.89	0.11	0.48	1.3
20	OPPR_8_20	0	-1.77	-0.21	0.74	1.24
21	OPPR_9_21	0	-0.6	-0.17	0.35	0.42
22	OPPR_10_R_22	0	-2.07	-1.25	0.63	2.69
23	OPPR_11_23	0	-1.35	0.03	0.58	0.74
24	OPPR_12_24	0	-1.41	-0.18	0.48	1.11
25	COMM_1_25	0	-2.15	-0.21	2.37	
26	COMM_2_26	0	-1.64	-0.17	1.81	
27	COMM_3_27	0	-1.77	-0.5	2.28	
28	COMM_4_28	0	-2.08	0	2.08	
29	COMM_5_29	0	-2.33	-0.17	2.51	
30	COMM_6_30	0	-1.75	-0.29	2.04	
31	COMM_7_31	0	-1.27	-0.39	1.66	
32	COMM_8_32	0	-1.2	-0.36	1.56	
33	COMM_9_33	0	-2.02	-0.26	2.28	
34	COMM_10_34	0	-1.9	-0.28	2.18	

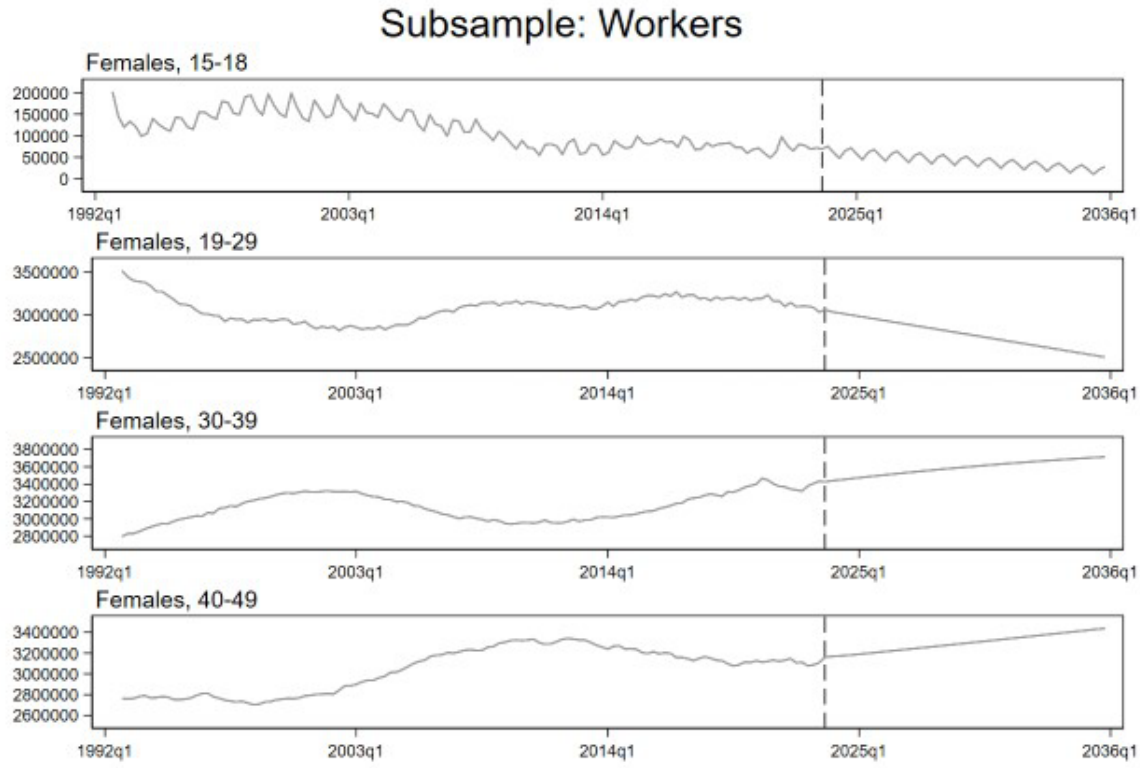
35	COMM_11_35	0	-2.31	-0.1	2.41	
36	COMM_12_36	DELETED				
37	COCO_1_37	DELETED				
38	COCO_2_38	0	-2.59	-0.09	2.68	
39	COCO_3_R_39	0	-1.62	-0.48	-0.36	2.46
40	COCO_4_40	0	-2.17	-0.46	2.63	
41	COCO_5_41	0	-2.46	0.06	2.39	
42	COCO_6_42	0	-3.08	0.14	2.94	
43	COCO_7_43	0	-2.37	-0.06	2.43	
44	COCO_8_44	DELETED				
45	COCO_9_45	DELETED				
46	COCO_10_46	0	-2	-0.61	0.3	2.31
47	COCO_11_R_47	0	-1.02	-0.72	-0.25	1.99
48	COCO_12_48	DELETED				
49	TCRE_1_49	0	-2.05	-0.07	0.99	1.12
50	TCRE_2_50	DELETED				
51	TCRE_3_51	DELETED				
52	TCRE_4_52	0	-1.39	-0.18	0.77	0.79
53	TCRE_5_53	DELETED				
54	TCRE_6_54	0	-1.48	-0.01	0.71	0.79
55	TCRE_7_55	0	-2.35	-0.11	1.03	1.42
56	TCRE_8_56	0	-2.32	-0.03	0.95	1.4
57	TCRE_9_57	DELETED				
58	TCRE_10_58	DELETED				
59	TCRE_11_59	DELETED				
60	TCRE_12_R_60	0	-3.66	-0.37	4.03	
61	INLI_1_61	0	-1.91	-0.06	1.97	
62	INLI_2_62	DELETED				
63	INLI_3_63	0	-2.3	-0.23	2.53	
64	INLI_4_64	0	-2.28	-0.09	2.36	
65	INLI_5_65	DELETED				
66	INLI_6_66	0	-2.92	0.28	2.64	
67	INLI_7_67	DELETED				
68	INLI_8_68	0	-2.31	0.06	2.24	
69	INLI_9_69	0	-3.13	0.28	2.85	
70	INLI_10_70	0	-2.22	-0.02	2.24	
71	INLI_11_71	0	-1.95	-0.1	2.05	

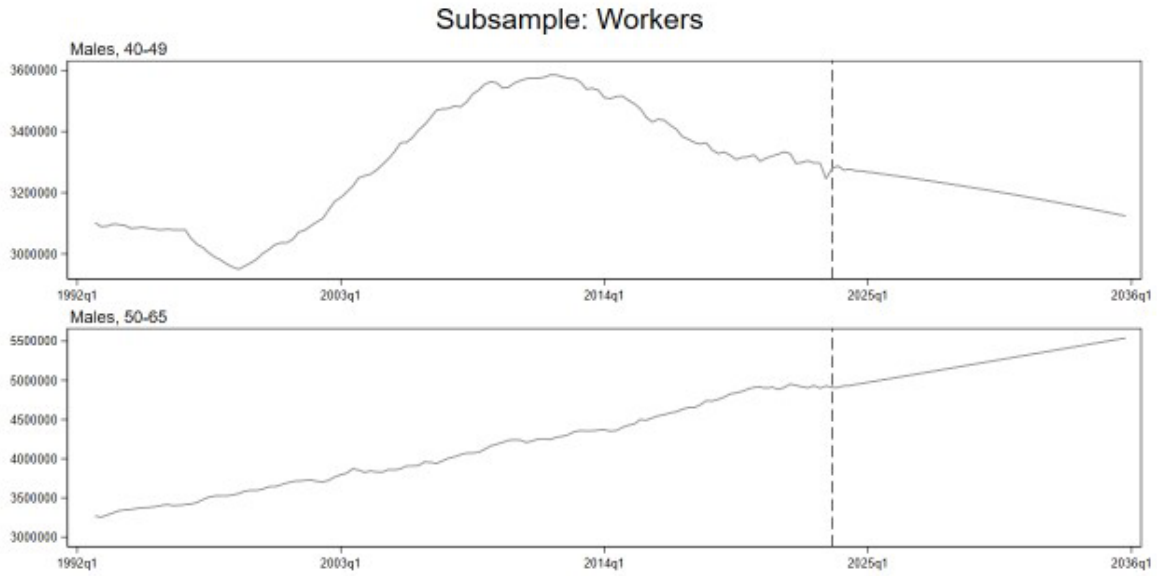
72	INLI_12_72	0	-2.15	-0.13	2.27	
73	O*NET_PSDM_73	0	-1.32	-0.63	0.24	1.72
74	O*NET_OPPR_74	0	-1.05	-0.81	0.25	1.61
75	O*NET_COMM_75	0	-1.35	-0.47	0.37	1.45
76	O*NET_COCO_76	0	-1.64	-0.7	0.75	1.58
77	O*NET_TCRE_77	0	-1.51	-0.4	0.34	1.57
78	O*NET_INLI_78	0	-1.68	-0.85	0.48	2.05
79	BIG_5_11_R_79	0	-1.85	-0.24	-0.01	2.11
80	BIG_5_21_80	0	-2.04	-0.49	0.19	2.35
81	BIG_5_31_81	0	-2.91	-0.55	0.59	2.87
82	BIG_5_12_82	0	-3.41	0.48	2.93	
83	BIG_5_22_R_83	0	-2.79	0.46	0.25	2.08
84	BIG_5_32_84	0	-1.69	-0.53	2.22	
85	BIG_5_13_R_85	0	-2.06	0.13	-0.06	2
86	BIG_5_23_R_86	0	-2.31	0.09	-0.14	2.36
87	BIG_5_33_87	0	-1.49	1.49		
88	BIG_5_14_R_88	0	-1.67	-0.22	-0.09	1.99
89	BIG_5_24_R_89	0	-1.8	-0.06	0.15	1.71
90	BIG_5_34_90	0	-0.96	-0.57	1.53	
91	BIG_5_15_91	0	-1.76	-0.24	0.3	1.7
92	BIG_5_25_R_92	0	-2.61	-0.42	0.59	2.44
93	BIG_5_35_93	0	-1.78	-0.32	2.1	

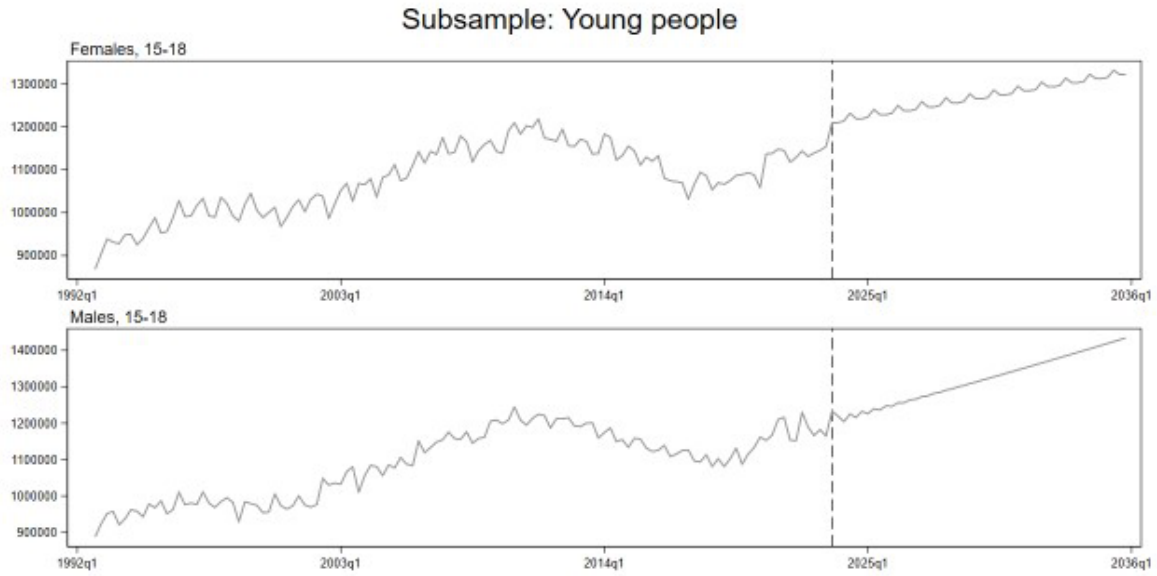
Appendix F Projected population totals through to 2035, for each of the variables and combinations of variables used in the re-weighting process

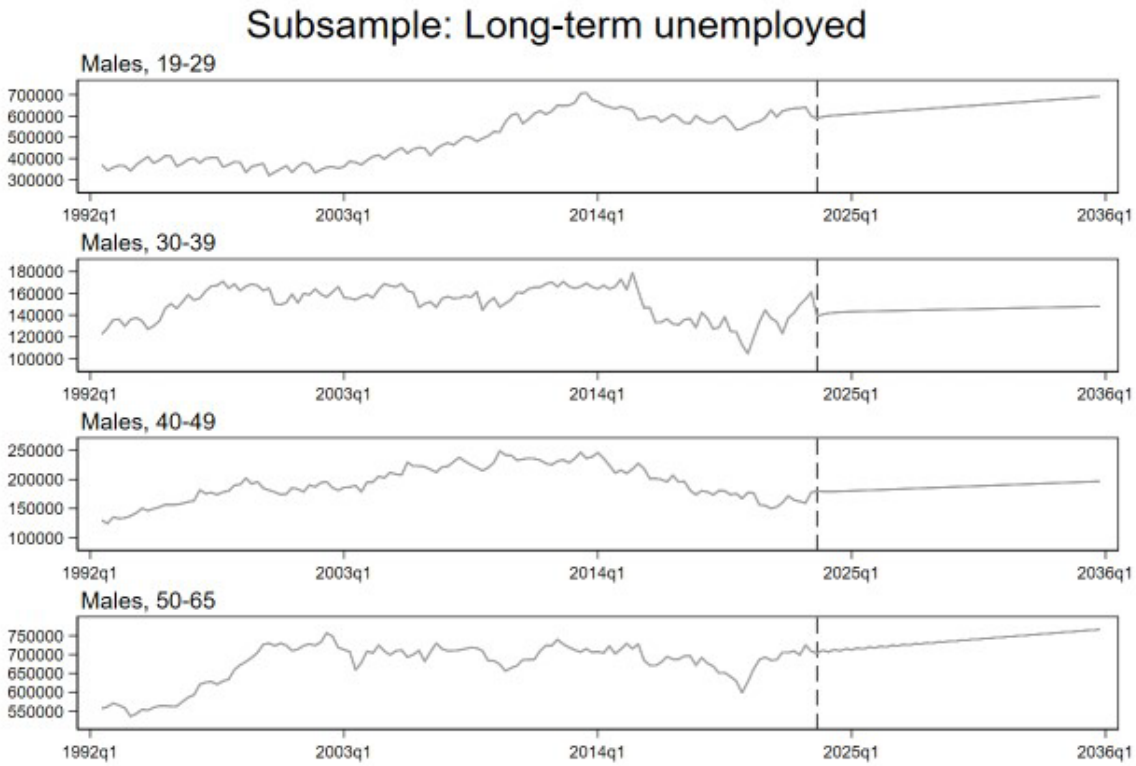
Panel A: Age and sex combination

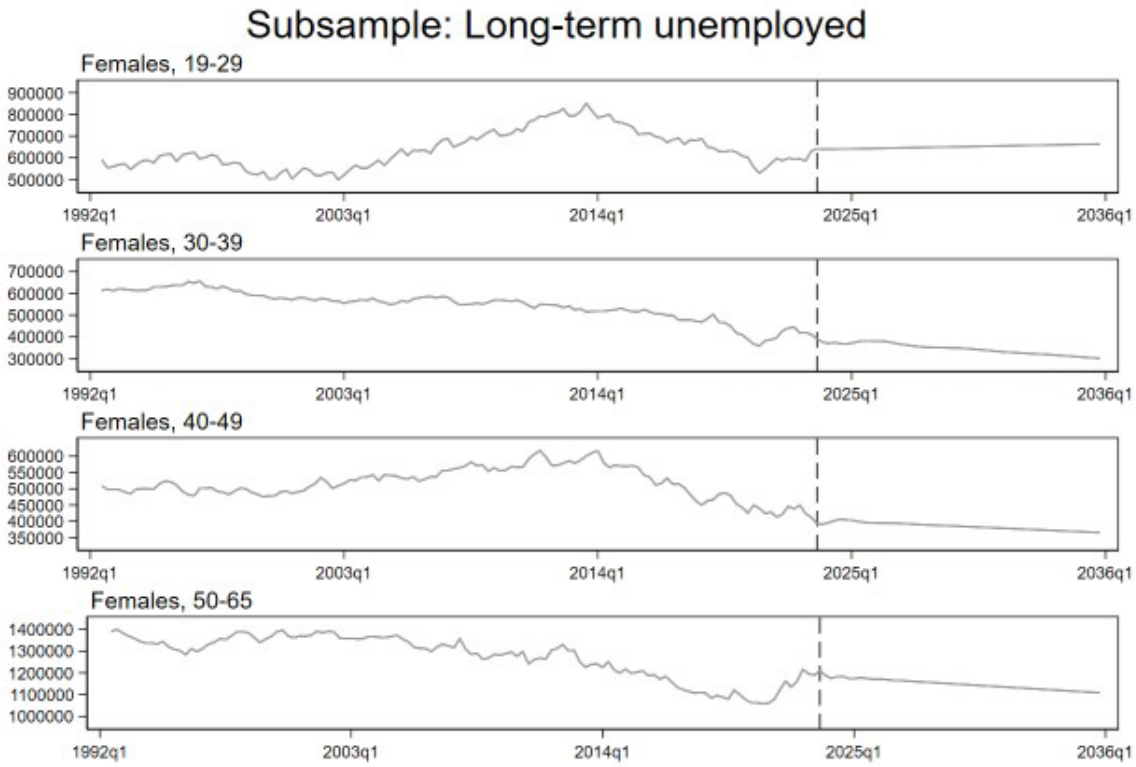




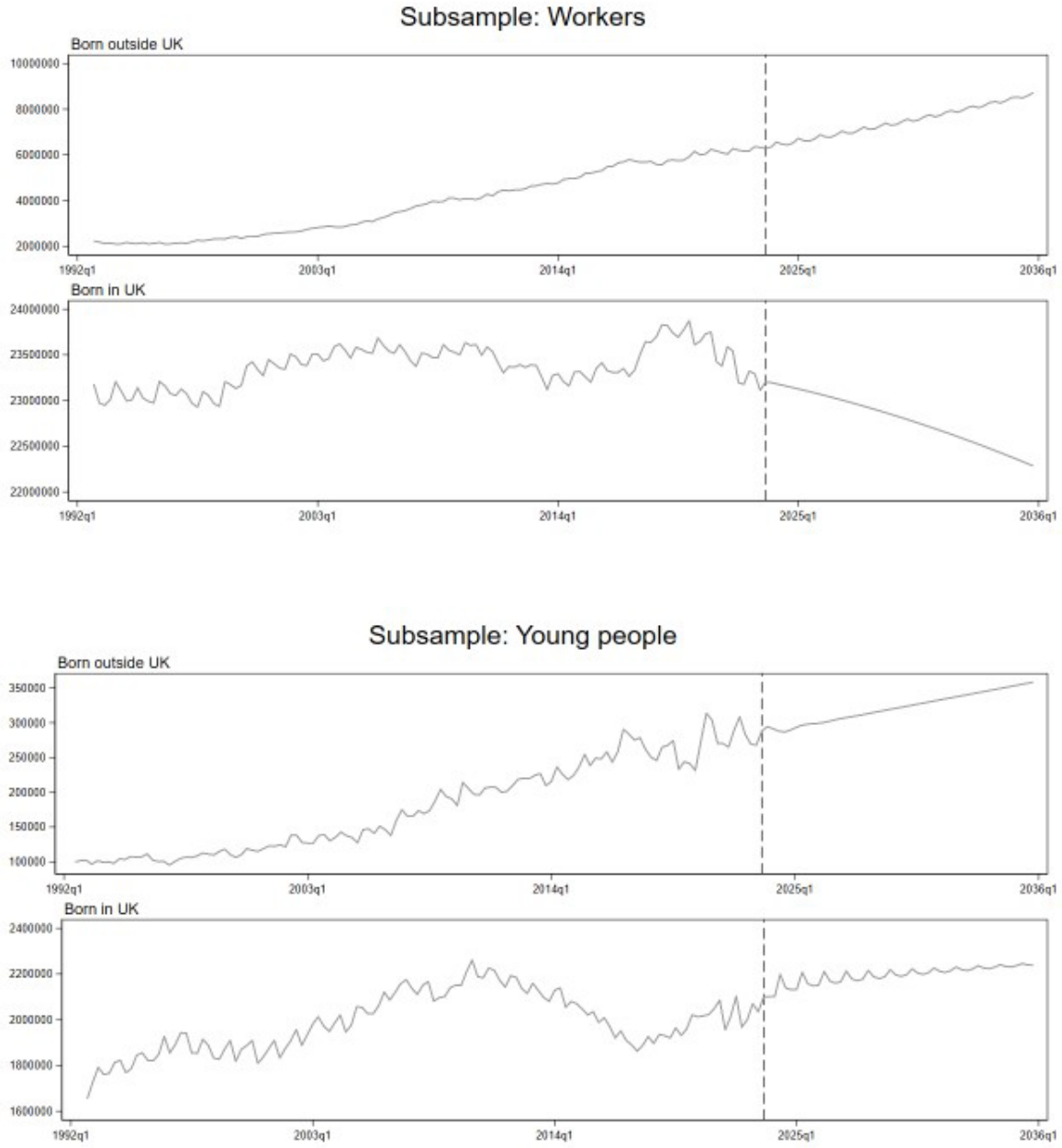

Subsample: Workers
Males, 40-49
3600000
3400000
3200000
3000000
1992q1
2003q1
2014q1
2025q1
2036q1
Males, 50-65
5500000
5000000
4500000
4000000
3500000
3000000
1992q1
2003q1
2014q1
2025q1
2036q1

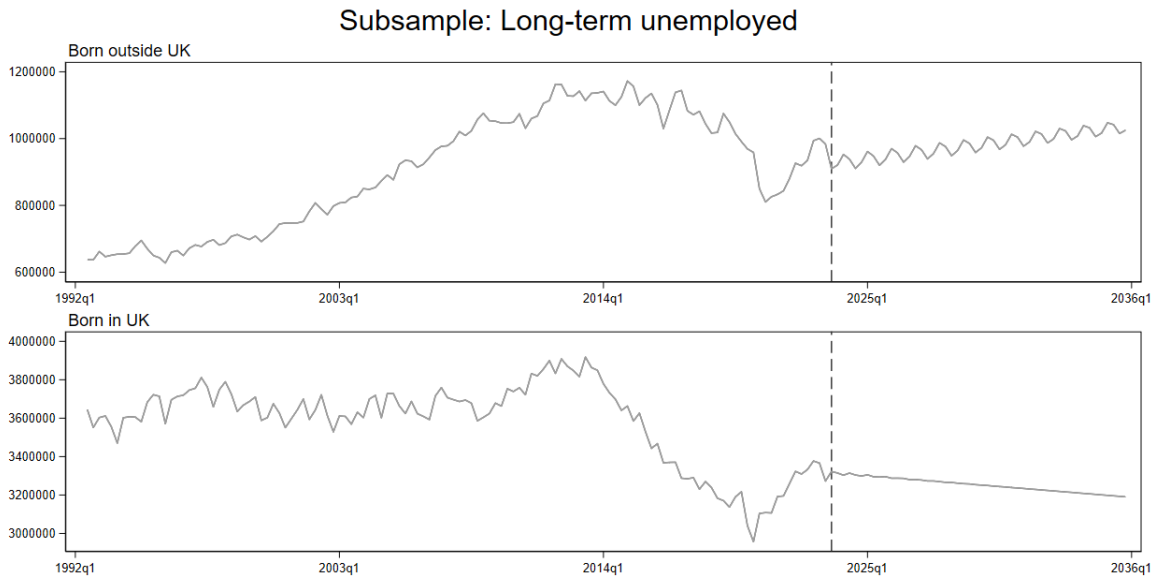

Subsample: Young people
Females, 15-18
1300000
1200000
1100000
1000000
900000
1992q1
2003q1
2014q1
2025q1
2036q1
Males, 15-18
1400000
1300000
1200000
1100000
1000000
900000
1992q1
2003q1
2014q1
2025q1
2036q1


Subsample: Long-term unemployed
Males, 19-29
700000
600000
500000
400000
300000
Males, 30-39
180000
160000
140000
120000
100000
Males, 40-49
250000
200000
150000
100000
Males, 50-65
750000
700000
650000
600000
550000
1992q1
2003q1
2014q1
2025q1
2036q1

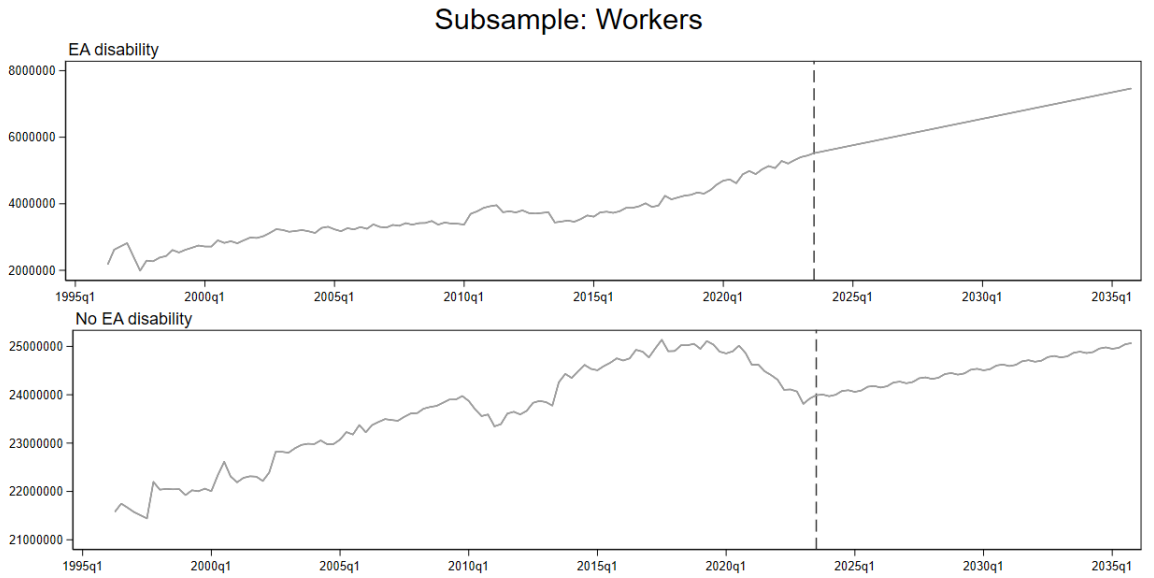

Subsample: Long-term unemployed
Females, 19-29
900000
800000
700000
600000
500000
Females, 30-39
700000
600000
500000
400000
300000
Females, 40-49
600000
550000
500000
450000
400000
350000
Females, 50-65
1400000
1300000
1200000
1100000
1000000
1992q1
2003q1
2014q1
2025q1
2036q1

Panel B: Country Of Birth

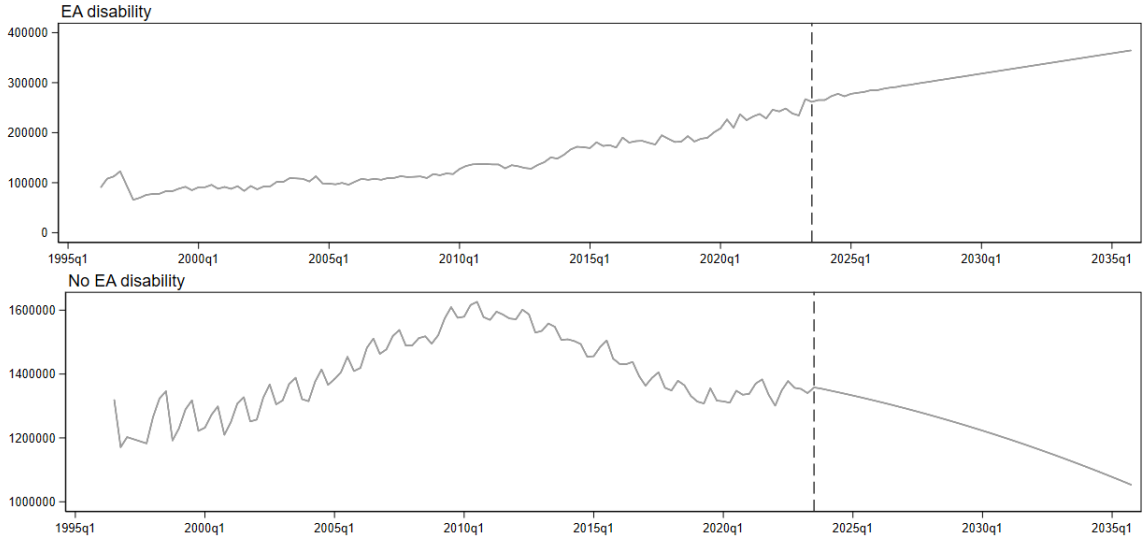




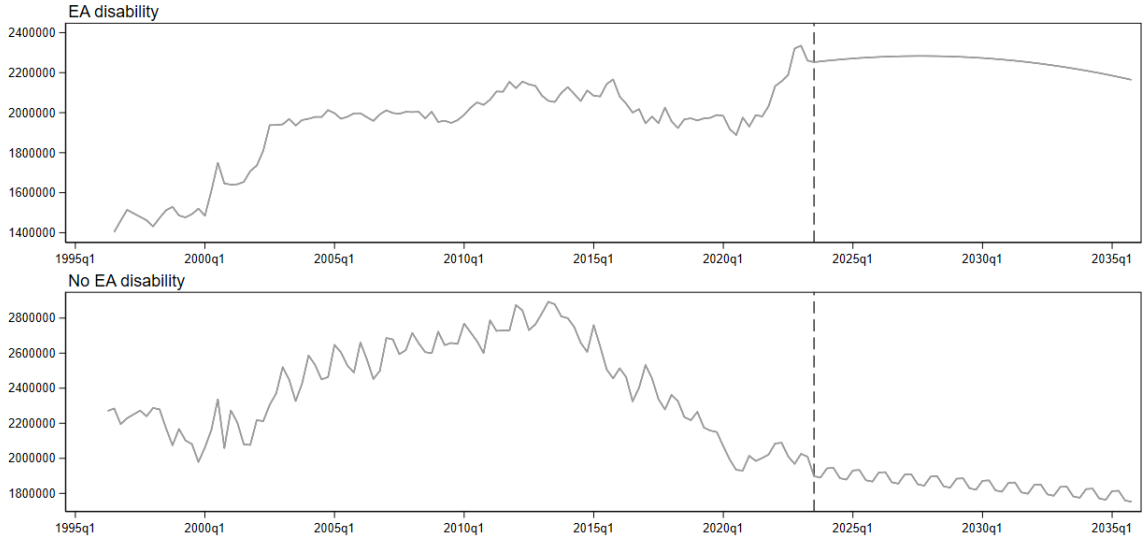
Panel C: Disability



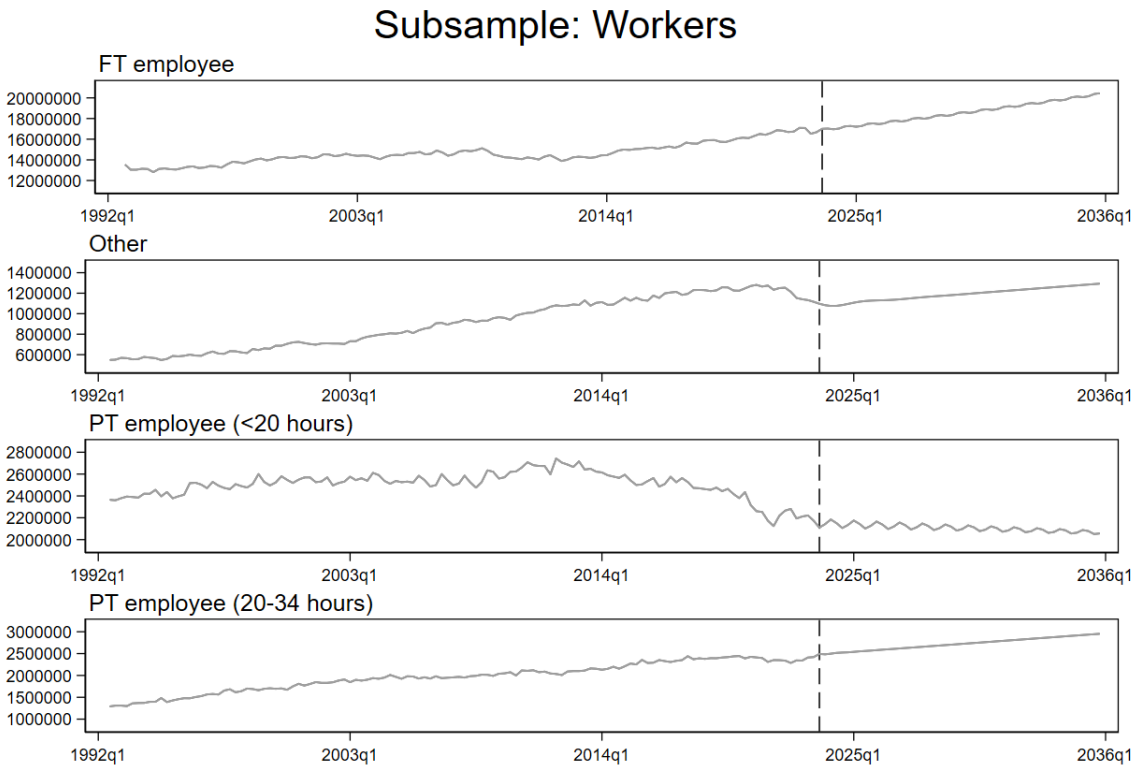
Subsample: Young people

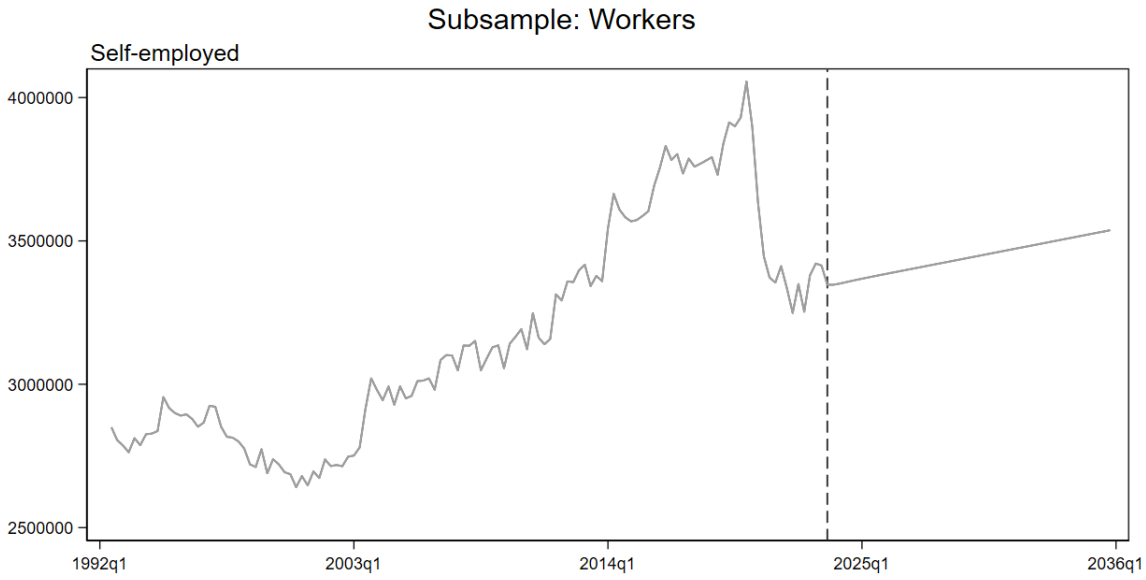


Subsample: Long-term unemployed

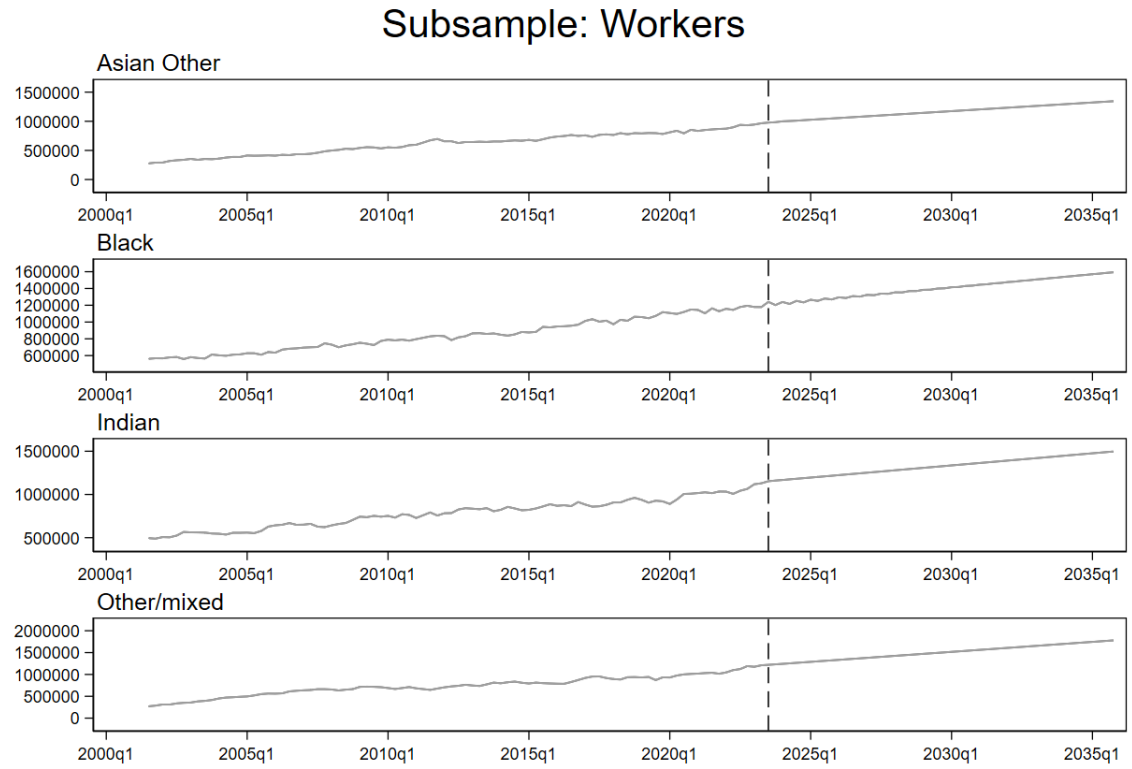


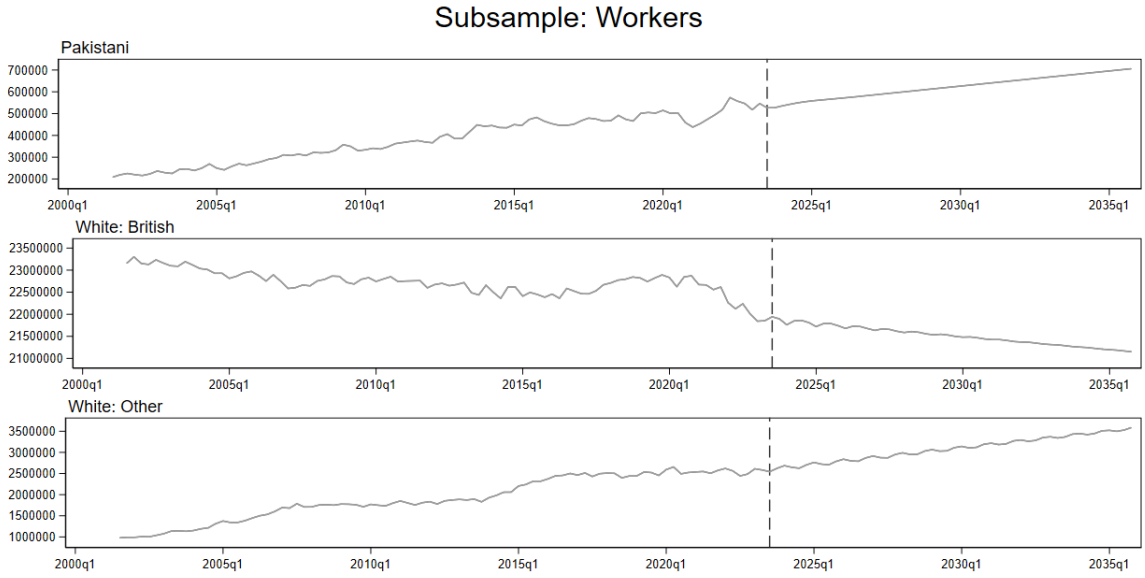
Panel D: Employment status

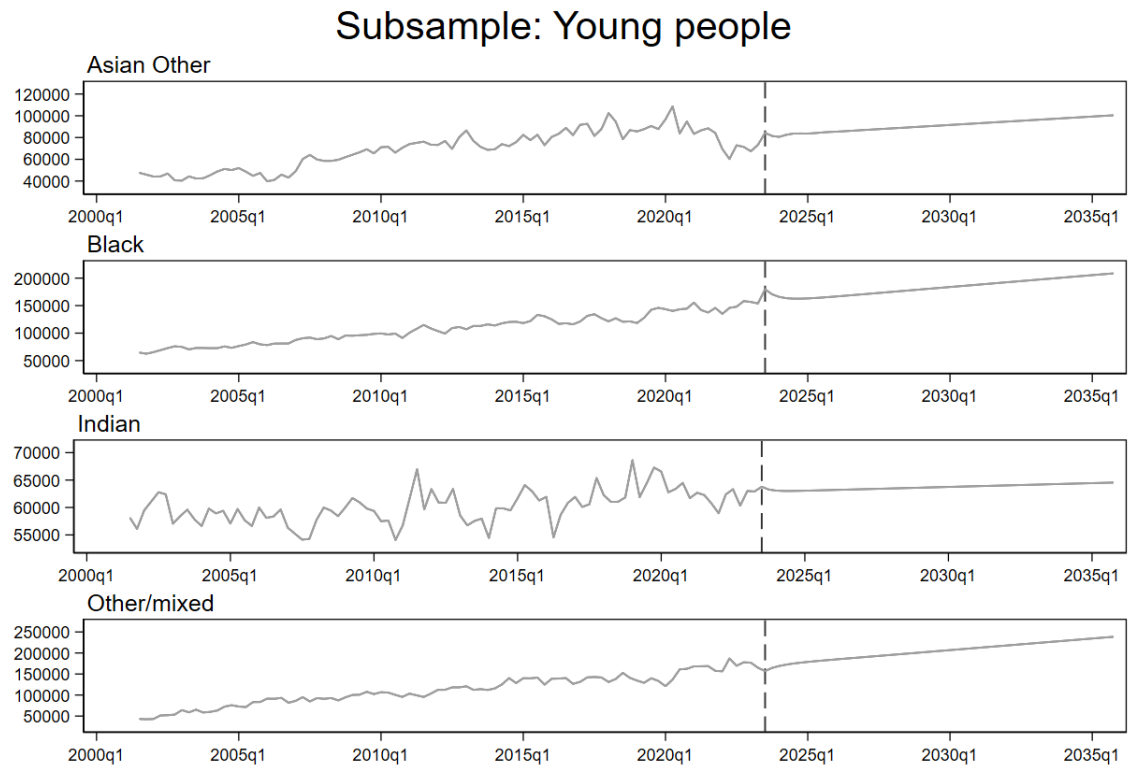


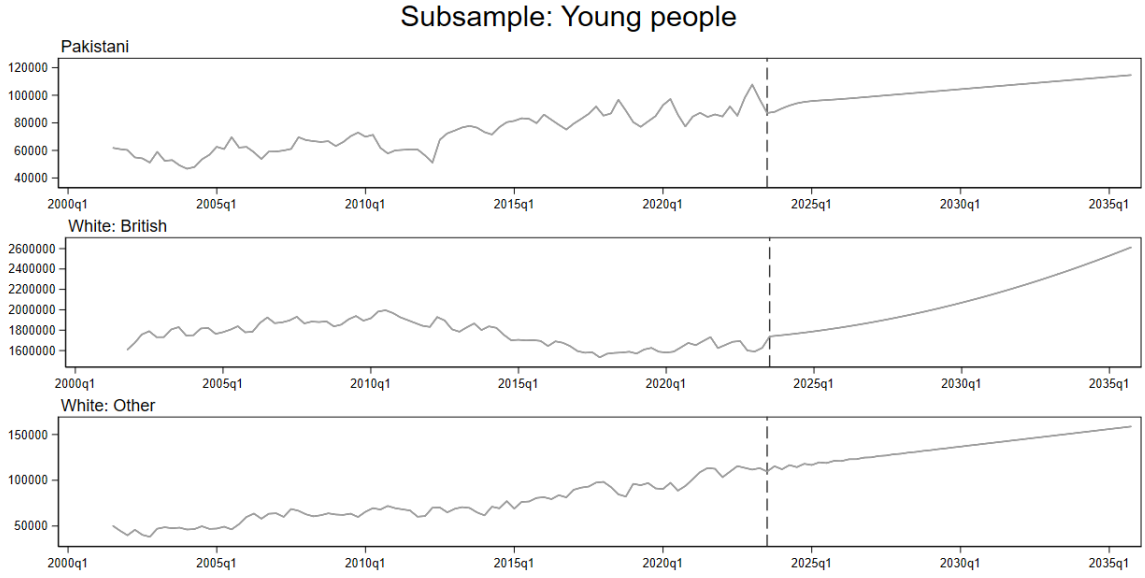


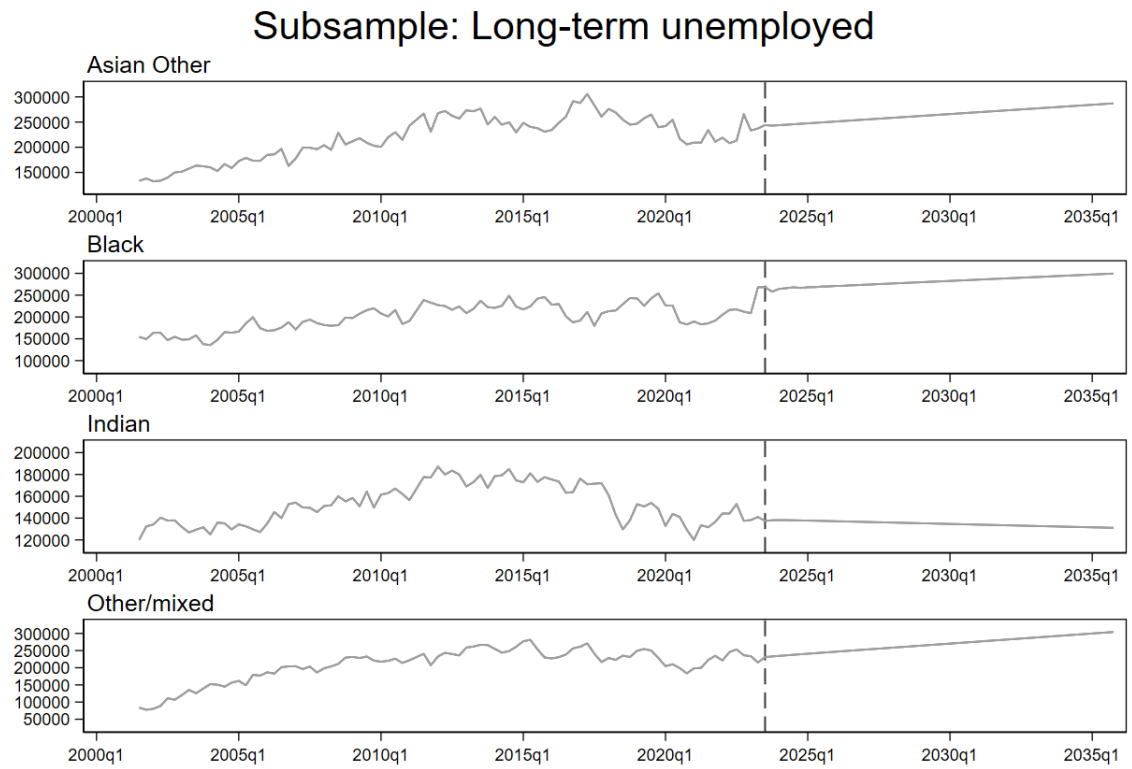
Panel E: Ethnicity

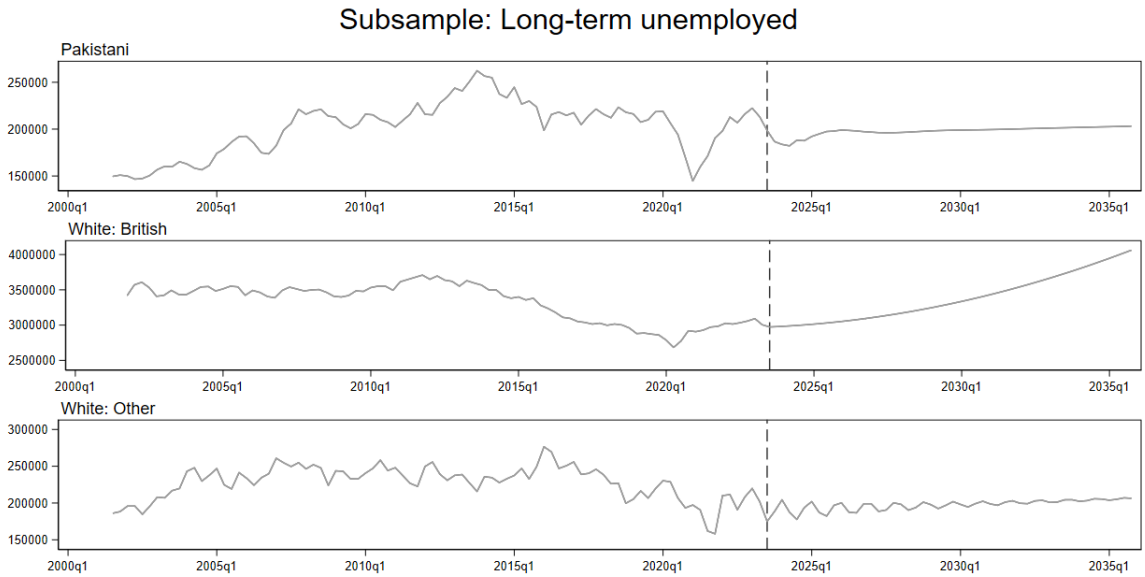




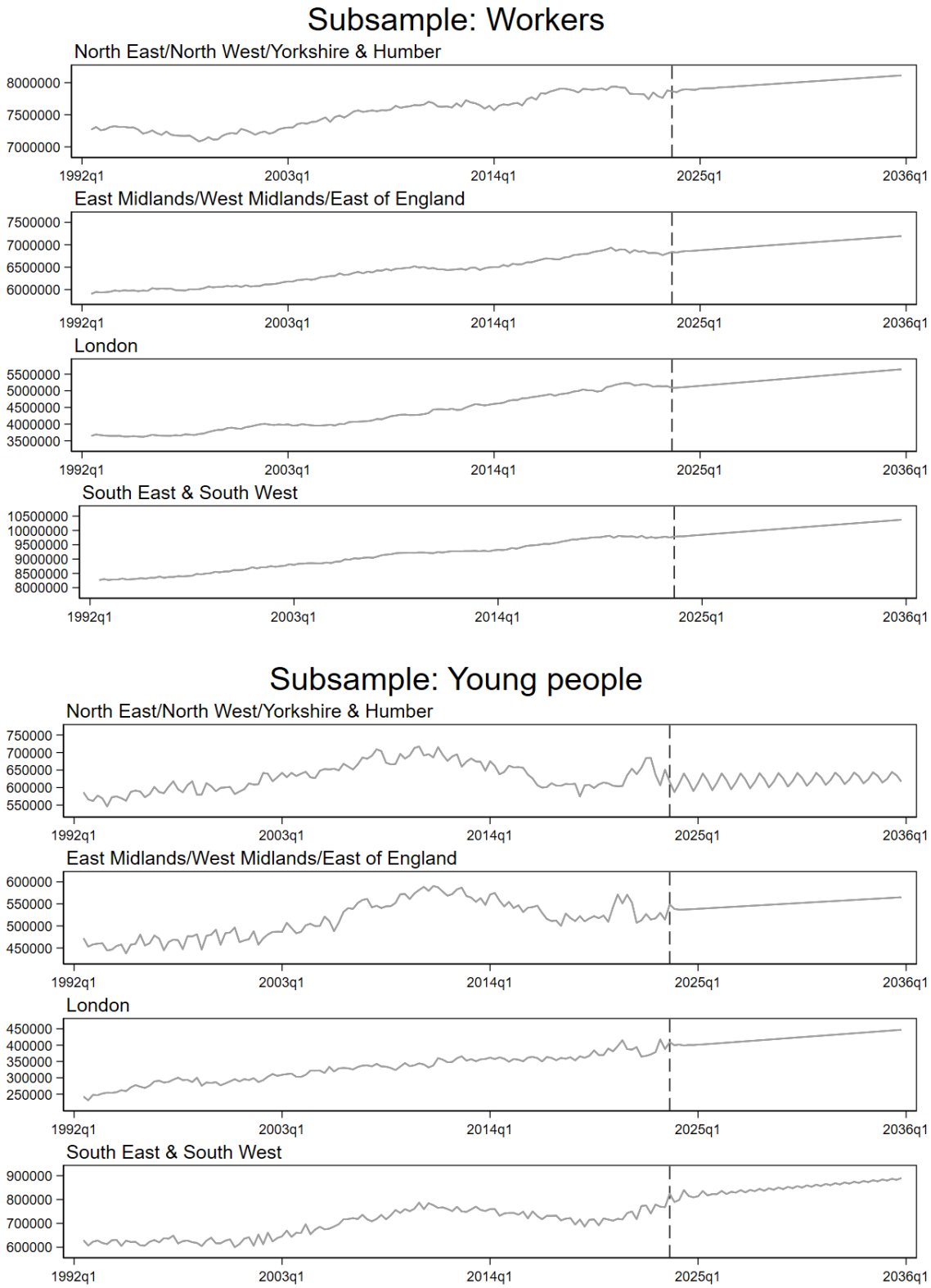

Subsample: Young people
Asian Other
120000
100000
80000
60000
40000
2000q1
2005q1
2010q1
2015q1
2020q1
2025q1
2030q1
2035q1
Black
200000
150000
100000
50000
Indian
70000
65000
60000
55000
Other/mixed
250000
200000
150000
100000
50000

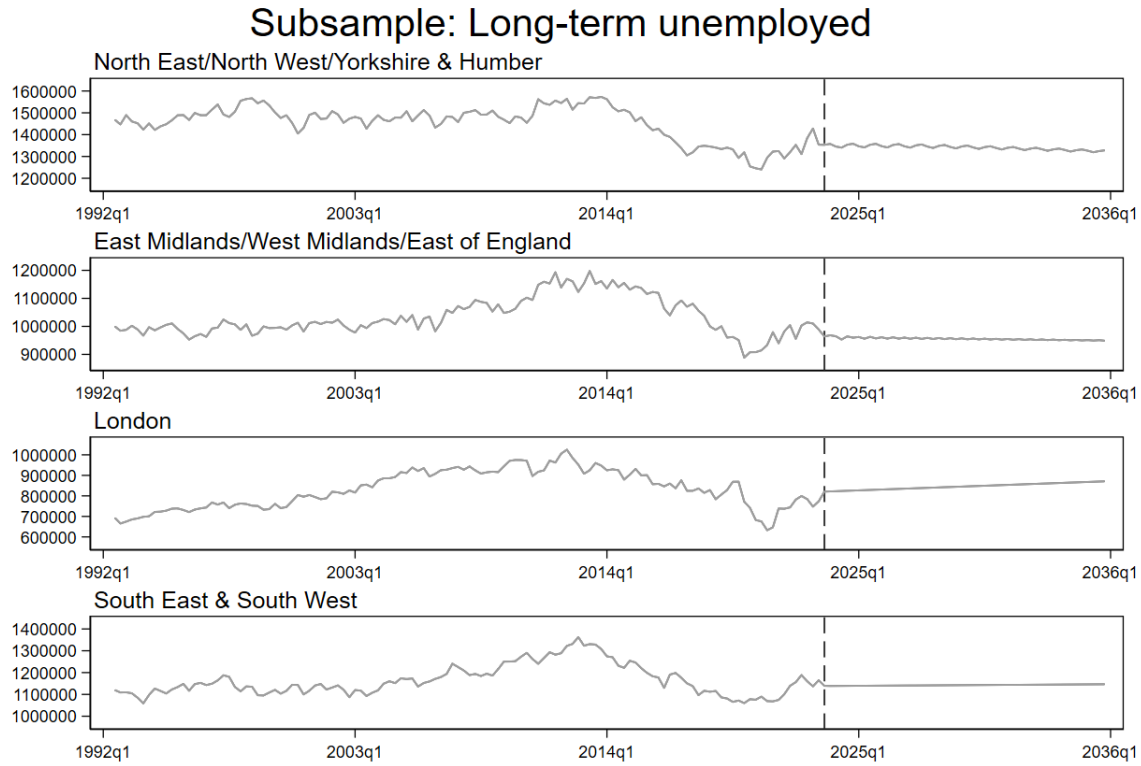

Subsample: Young people
Pakistani
120000
100000
80000
60000
40000
2000q1
2005q1
2010q1
2015q1
2020q1
2025q1
2030q1
2035q1
White: British
2600000
2400000
2200000
2000000
1800000
1600000
White: Other
150000
100000
50000


Subsample: Long-term unemployed
Asian Other
300000
250000
200000
150000
2000q1
2005q1
2010q1
2015q1
2020q1
2025q1
2030q1
2035q1
Black
300000
250000
200000
150000
100000
Indian
200000
180000
160000
140000
120000
Other/mixed
300000
250000
200000
150000
100000
50000

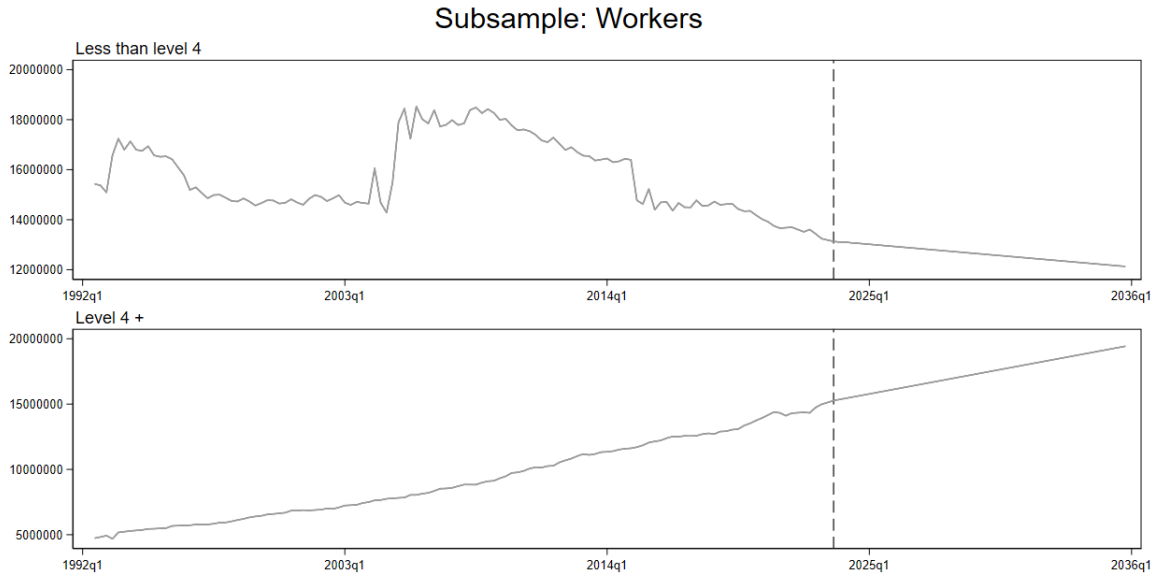

Subsample: Long-term unemployed
Pakistani
250000
200000
150000
White: British
4000000
3500000
3000000
2500000
White: Other
300000
250000
200000
150000

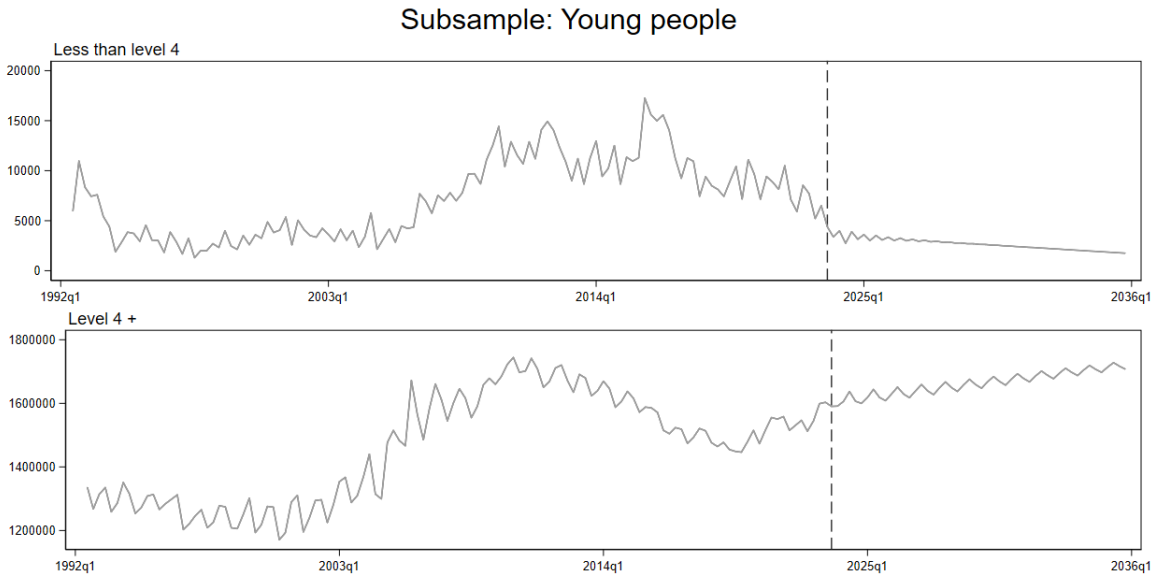
Panel F: Geography

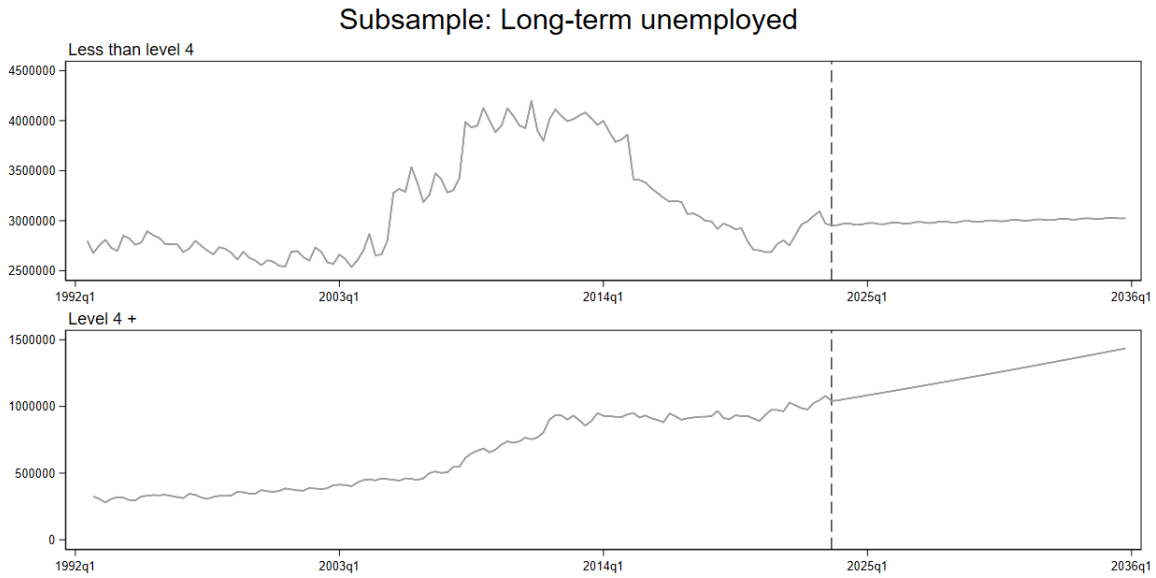




Panel G: Qualification




Subsample: Young people
Less than level 4
20000
15000
10000
5000
0
1992q1
2003q1
2014q1
2025q1
2036q1
Level 4 +
1800000
1600000
1400000
1200000
1992q1
2003q1
2014q1
2025q1
2036q1


Subsample: Long-term unemployed
Less than level 4
4500000
4000000
3500000
3000000
2500000
1992q1
2003q1
2014q1
2025q1
2036q1
Level 4 +
1500000
1000000
500000
0
1992q1
2003q1
2014q1
2025q1
2036q1

The Mere, Upton Park, Slough, Berks SL1 2DQ
T: +44 (0)1753 574123
F: +44 (0)1753 691632
enquiries@nfer.ac.uk

www.nfer.ac.uk

The Skills Imperative 2035