What Drives Reskilling Decisions? Evidence from a Discrete Choice Experiment with Unemployed Jobseekers

Pre-Analysis Plan

Benjamine Dejardin, Alexia Delfino, Elisabeth Leduc, Raffaella Sadun, Ilan Tojerow

November 14, 2024

1 Introduction

Rapidly evolving labour markets have led to deteriorating employment opportunities for workers with few or obsolete skills, and increasing labour and skills shortages for employers. This has led to a growing mismatch between the available jobs and the qualifications of workers seeking employment. As a result, reskilling the workforce is becoming an essential strategy to address these challenges.

Our project investigates whether jobseekers are willing to reskill and, if so, what this decision depends on. We will explore this question with a discrete choice experiment implemented with the Public Employment Service (PES) operating in Wallonia, Belgium. Using responses from approx. 3,000 Belgian jobseekers, we will explore their willingness to enroll in demand-driven occupational training programs, assess their willingness-to-pay (WTP) for different training features, and compare these preferences with those for job-related characteristics. Our study will thus focus on the importance of key policy-relevant factors, as well as the importance of working conditions in the target occupation, in explaining the decision to reskill. Moreover, we will examine whether jobseekers' willigness to reskill (and their WTP for different training and job features) depends on: their interest in and beliefs about the target occupation, the distance between their target occupation and previous work experience, as well as other personal characteristics. By linking survey data with administrative records, we will be able to relate jobseekers' stated preferences to their actual training decisions.

2 Experimental Design

Observational data on training take-up reflects decisions made under specific conditions, making it difficult to assess the impact of individual factors on decision-making. Our discrete choice experiment (DCE) allows to circumvent this issue, by asking participants to make hypothetical decisions between different training options, where the characteristics of the training options are observed and can be experimentally manipulated. In particular, we aim to present respondents with training options which differ in the extent to which they enable trainees to enter an occupation in high-demand versus occupations that they have a preference for. This allows us to understand how jobseekers trade-off personal taste with economic opportunities in training choices.

We aim to get approximately 3,000 responses. Data will be collected primarily through emails sent by the PES in late November and/or early December 2024. At a further stage, more data might be collected in person at PES offices in order to increase sample size and diversify selection into the survey.

Our target population of respondents consists of unemployed jobseekers registered with the Walloon Public Employment Service (PES) for at least two weeks, who are aged under 55, and who are not exempt of searching for a job. The survey, which takes approximately 10 minutes to complete, is described in more detail in the following sections.

2.1 Structure of the survey

2.1.1 Occupational preferences

The first section of the survey aims at defining the occupational preferences of respondents in order to offer them training options that match these preferences. To this end, participants are first asked to choose the sector in which they would most like to find employment.

Next, respondents are presented with a list of ten occupations within their preferred sector, and are asked to choose their preferred option (named "occupation ranked #1" henceforth). We then show respondents a subset of occupations from which they should choose once more (named "occupation ranked #2" henceforth). This list of occupations excludes the favourite one as well as any occupations which are not in shortage. A randomly selected half of respondents are informed that the second list includes only shortage occupations. This creates random variation in the conditions under which individuals will make their training decisions in the following sections.

The training options proposed in the following sections of the survey will be anchored in the occupational preferences elicited in this first section. By doing so, our aim is to mimic as closely as possible the real-life decision process of choosing a training. It is indeed likely that jobseekers would be interested in enrolling in trainings related to more than one occupation. As such, when choosing a training program, they are likely to make trade-offs between training options that target two different occupations, depending on training- or job-related characteristics (e.g., level of employer demand, length of the training, etc.). We discuss this in more detail in the following sections.

2.1.2 Training decisions under different scenarios: training features vary

In the second part of the survey (after detailing their occupational preferences), respondents are asked to choose their preferred option between:

- Training A: a training targeting occupation ranked #1
- Training B: a training targeting occupation ranked #2
- No training

Participants make this choice five consecutive times, with each scenario featuring different combinations of training characteristics. Table 1 shows the list of training characteristics that we consider in our experiment, as well as the levels each feature can take. In the survey, we randomize the order in which training features are presented.

Table 1: Training features and levels

Subject	Occupation ranked #1 Occupation ranked #2
Financial compensation (in addition to UI benefits)	304€/month + 0€[500€]2000€ at end of training
Length	1 month 4 months 9 months
Employer involvement	Yes No
Format	In person (group) Online (alone) Hybrid (mix of online and in person)
Travel Time to Training Center	20 min 40 min 60 min
Waiting Time	Immediate 6 months

We inform respondents about the meaning of each training characteristic at the start of this survey section. Moreover, in each choice scenario, they can see the definition of each feature by hovering or clicking on the feature's name. Training features are defined as follows:

- Subject: Occupation for which the training prepares you.
- Financial compensation (in addition to UI benefits): Monthly compensation for being enrolled in a training + financial bonus received in case of successful completion of the training.
- Length: Number of months during which you would be taking the course.
- Employer involvement: Training developed with an employer who is looking to hire certain profiles and who commits to hiring 80% of the trainees.
- Format: Training delivered online, face-to-face or a mixture of both. Online training is done alone, while face-to-face training is delivered in groups.
- Travel time: Travel time between your home and the training centre
- Waiting time: Waiting time before starting the training.

The choice scenarios with the combination of these training characteristics are built using a D-efficient design to maximize the efficiency of the data collected (through the stata code "d-create"). This method maximises the D-efficiency of the design based on the variance-covariance matrix of the parameter estimates of the conditional logit model (Cook and Nachtsheim, 1980; Zwerina et al., 1996; Carlsson and Martinsson, 2003). This method is employed to avoid using a full factorial design, i.e. a design in which each combination of the proposed factors is presented as an alternative to the respondent. The D-efficient design allows a precise identification of the parameters of interest, without imposing huge cognitive burden to the respondents.

In total, we select 15 scenarios which are divided into 3 blocks, with each respondent being assigned to one block (and thus seeing 5 scenarios). As an example, Figure 1 shows a caption of one of the scenarios shown to respondents.

The choices made in our survey are hypothetical, and we thus need take several steps to try and maximize the chances that respondents make choices that would match as closely as possible their behaviour in reality. First, we remind participants that there are no right or wrong answers, and that they can also choose neither of the training options. Second, we inform them about the anonymity of their responses, so that they do not fear that the PES might know what they said.

In addition, to further encourage honest answers, we set up an incentivization for truthful revelation. Specifically, at the start of the survey, participants are informed that, based on

Figure 1: Example of choice scenario - Varying training features

	Formation A: Accueillant(e) d'enfant	Formation B: Aide familiale
Implication employeur	Oui	Non
Temps de trajet	20 minutes	40 minutes
Délais d'attente	Immédiat	6 mois
Format	Mixte (en présentiel et en ligne)	En présentiel (en groupe)
Compensation financière (en plus de vos allocations habituelles)	304€/mois	304€/mois + 2000 € en fin de formation
Durée	9 mois	4 mois

○ Formation A

○ Formation B

○ Aucune des deux

→

their answers, they will be provided with useful information about jobs or trainings that seem to be of interest to them. They will then be able to use this information for their job or training search, and/or during their next meeting with a caseworker. At the end of the survey, we will ask respondents how much they care about this information to check the strength of this incentivization (more details in section 2.1.5).

Finally, we will have access to administrative records which provide information on all training activities actually performed. This will allow us to check to what extent the hypothetical choices made in our survey match actual training choices in real life. Moreover, we will perform exploratory analyses to check whether respondents change their training choices after receiving information about occupation #2 being in shortage. Note, however, that with 3,000 respondents we will likely be underpowered to detect statistically significant effects on training behaviour.

2.1.3 Perceptions and beliefs

We are also interested in understanding to what extent training preferences relate to beliefs about the occupation a training targets. Therefore, in the third part of our survey, we include a series of questions aiming to understand how respondents perceive the occupations they ranked #1 and #2. Specifically, we ask them about their:

- Level of interest in the two occupations (1-100 scale);
- Beliefs about working conditions: overall, salary, career evolution prospects, work demands, schedule flexibility, work-life balance (5-point scale);

- Perceived level of employer demand for the occupation (5-point scale);
- Beliefs on the likelihood of being hired in the occupation (1-100 scale) with and without following a training program

2.1.4 Training decisions under different scenarios: job characteristics vary

When making reskilling decisions, jobseekers probably consider the working conditions in the job that the training would prepare them for. In other words, in addition to the characteristics of the training, job-related factors likely play a role in explaining why jobseekers are willing (or not) to reskill. In the fourth part of the survey, we are interested in exploring the importance of these job-related characteristics in the decision to reskill. Our aim here is to compare the importance of "policy relevant" factors (i.e., training design), with "labour market" factors related to working conditions in the target occupations.

To do so, we repeat the training choice exercise, but focusing on the role of job-related characteristics. Specifically, we will show respondents two training options that both target occupation ranked #2. In our context, the variation in job characteristics will come from the fact that the training options are organized with the involvement of employers who differ in terms of the job characteristics they offer to their workers. Since employers who are involved in the design and organization of trainings with the PES typically commit to hiring about 80% of trainees, we can induce random variation in the characteristics of the jobs to which each training option would likely lead.

Participants can then choose whether they would enrol in:

- Training A: organized with an employer offering a given set of job characteristics
- Training B: organized with an employer offering a different set of job characteristics
- Neither

Table 2 shows the list of job characteristics that we consider in this section of our experiment, as well as the levels each feature can take. Respondents are informed about the meaning of each characteristic at the start of this survey section. Moreover, in each choice scenario, they can see the definition of each feature by hovering on its name. The definition of each feature is the following:

- Salary: The monthly salary offered by this employer compared to his sector (above, equal to, or below sector average).

- Career evolution prospects: The extent to which the employer offers options to change jobs horizontally (same level of responsibility) or vertically (promotion to higher level of responsibility).

- Work demands: Whether this employer offers jobs that expose workers to occupational risks that have an impact on physical or mental health (e.g. physical workload, work organisation, safety risks or emotional workload). Although work demands vary across occupations, they can also vary among employers within a given occupation, as some employers provide relatively more protection against these hazards.

- Schedule: the extent to which workers can decide on the hours they work, or whether they are imposed by the employer. This includes items related to: arrival and departure times, teleworking, the choice between full-time or part-time work, but also how predictable the work schedule is.

Table 2: Job characteristics and levels

Training characteristics	
Financial compensation (in addition to UI benefits)	304€/month + 0€[500€]2000€ at the end of the training
Length	1 month 4 months 9 months
Job-related characteristics	
Salary	Below average Average Above average
Career evolution prospects	Limited Numerous
Work demands	Low Moderate High
Schedule	Schedule imposed by employer Worker involved in setting schedule

Note that Table 2 also includes two training-related characteristics: the financial compensation for doing the training, and the length of the training. We include these items in order to measure the WTP for job-related characteristics in the same unit as our WTP measures for training characteristics. We describe this in more detail in section 4.

As for the scenarios with varying training features, these scenarios are built using a D-efficient design. In this part, we select a total of 12 scenarios which are divided into 4 blocks, with each respondent being assigned to one block (and thus seeing 3 scenarios). As an example, Figure 2 shows a caption of one of the scenarios shown to respondents.

Figure 2: Example of choice scenario - Varying job-related features

	Formation A: Aide familiale	Formation B: Aide familiale
Caractéristiques de formation		
Durée	1 mois	4 mois
Compensation financière (en plus de vos allocations habituelles)	304€/mois + 500 € en fin de formation	304€/mois + 1500 € en fin de formation
Caractéristiques de l'employeur impliqué dans la formation		
Salaire mensuel	Supérieur à la moyenne du secteur	Inférieur à la moyenne du secteur
Possibilités d'évolution de carrière	Nombreuses	Limitées
Pénibilité	Faible	Elevée
Flexibilité des horaires	Horaire imposé par l'employeur	Travailleur impliqué dans la décision des horaires

○ Formation A

○ Formation B

○ Aucune des deux

→

2.1.5 Additional questions

Finally, at the end of the survey, we include questions relating to the following items that might explain the willingness to reskill:

- Number of children;
- Importance of accessibility of the training center by car and public transport;
- Importance of formal certification at the end of a training;
- Preference for doing a traineeship, and whether it should be organized throughout the training or at the end;
- How much respondents would like to find employment in the same occupation as the one they held in their previous job;
- What factor is the main driver / barrier for them when choosing to enrol in a training.

This information is not included in the administrative records and therefore can only be known if collected in the survey.

At the end of the survey, we ask a final question aiming to check the effectiveness of our incentivization for truthful revelation described in section 2.1.2. In particular, we ask respondents:

When you validate this survey, we will provide you with useful information for your job search. This information will be based on your answers to the survey. How interested are you in this information? Not interest – A little interest – Very interested.

This will enable us to check whether our estimates differ for respondents who care relatively more (i.e., are "a little interested" or "very interested") versus less (i.e., are "not interested") about the incentive to answer truthfully.

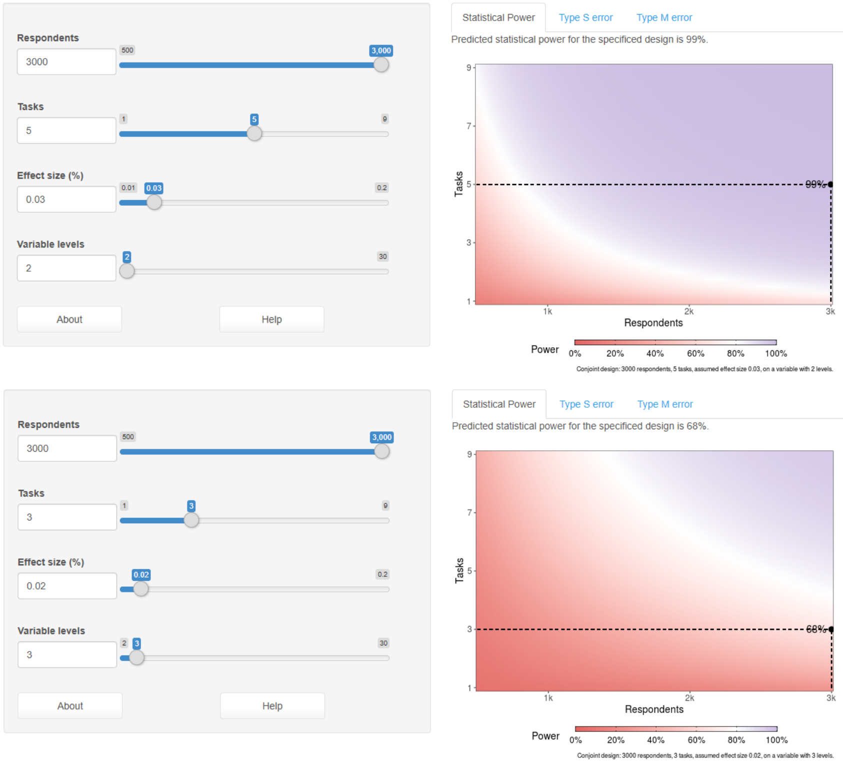
2.2 Power analysis

In discrete choice experiments, statistical power is influenced positively by the sample size and the number of "tasks" (choice sets) assigned to each respondent; it is influenced negatively by the number of alternatives for each attribute. We focus our analysis on the Average Component Marginal Effect (ACME), the most commonly used causal measure of interest that is used to analyse conjoint data. The ACME represents the causal effect of shifting one attribute of a conjoint profile on the probability of choosing that profile while averaging over the distribution of the other profile attributes. For example, in our setting the ACME represents the change in the probability of choosing one training when the training duration changes from one month to four months.

For our power analysis, we follow Stefanelli and Lukac (2020) and set the relevant parameters for our discrete choice experiment.[1] In their literature review, Stefanelli and Lukac (2020) find a median ACME equal to 0.05. In our setting, being able to detect an ACME equal to 0.05 means that we can detect when a change in one training attribute induces at least 5 percentage points increase in the probability of choosing such training, all else being equal. Given that about 25% of unemployed jobseekers enrol in a training in any given year, an ACME of 0.05 would imply a 20% proportional increase as a result of a change in an attribute from one level to another. We therefore analyze the required sample to be able to estimate comfortably (i.e. with a target statistical power of at least 80%) an ACME of 0.02-0.03. According to this exercise, and with between 2 and 3 levels per attribute, our target number of respondents is 3,000. Figure 3 illustrates our power analysis under different assumptions about expected effect size and variable levels.

[1] See their power simulation tool on: `https://mblukac.shinyapps.io/conjoints-power-shiny`

Figure 3: Power analysis



Notes: This figure shows a caption of our power analysis under different assumptions ($N = 3,000$; variable levels = 2 or 3); estimated effects = 0.02 or 0.03.

3 Data

We will combine data collected via our survey with data from administrative records. This will enable use to explore how preferences relate to various labour market history and personal characteristics.

From administrative records, we will have access to information on trainings followed (name of the training, start and finish date, reason for ending the training), job search behaviour (list of occupational preferences, number of connections to personal space on PES website), employment history (past occupations, employment spells), and personal characteristics (gender, age, nationality, district of residence, level of education, field of last degree).

We will receive administrative data for the full sample of respondents, as well as for a random sample of non-respondents. This will allow us to characterize respondents relative to the rest of the unemployed population. It will also allow us to show some descriptive statistics about email engagement (e.g., who was more likely to open the email and click on the survey).

From the survey, in addition to the training choices under different scenarios, we will have data on beliefs and perceptions about the target occupations (employer demand, level of interest, probability of being employed, working conditions), stated preferences for other training characteristics not included in the DCE (certification), as well as additional information which is not available in the administrative data (number of children, ease of transport, etc.).

4 Planned analyses

We plan on running a number of analyses, both of descriptive and causal nature. The ultimate aim of all these analyses is to understand which jobseekers could be induced to reskill and under what conditions.

4.1 Are jobseekers willing to follow training programs?

First, we will explore whether respondents display some willingness to reskill. We will do so by looking at the share of respondents that selects training A or B at least once, as well as the share of times each option is selected across scenarios.

4.2 What does the demand for training depend on?

Second, we will explore whether the willingness to reskill is correlated with observed characteristics such as demographics, level of interest in target occupations, perceptions and beliefs about working conditions in target occupations, demand of employers for target occupations, distance between target occupations and previous labour market experience, and level of interest in last occupation held. We will relate training choices with beliefs on target occupations both in levels and differences. For instance, we will compute the gap in perceived salary levels between occupation 1 and 2, and relate this to the willingness to choose training in occupation 1 vs 2 (ceteris paribus).

4.3 Do training program features matter for the decision to reskill?

First, we will look at the share of respondents who change their decision across different scenarios. This will already say something about whether training features matter *at all* in the decision to reskill.

Second, the key objective of this part of our analysis is to estimate the parameters (marginal utilities) of the discrete choice model through different logit specifications (Train, 2009), where the parameters of interest will be the features of each training (see section 2.1.2 for a complete list). The estimated model will include all training features, such that the choice is modelled as:

$$Y_{i,j} = \alpha_j + \beta_1 occupation_{i,j} + \beta_2 compensation_{i,j} + \beta_3 length_{i,j} + \beta_4 empinvol_{i,j} + \\ \beta_5 format_{i,j} + \beta_6 travelt_{i,j} + \beta_7 wait_{i,j} + X_i + \eta_{i,j}$$

Where $Y_{i,j}$ is the choice (training A, training B, or no training) made by individual i when presented with alternative j. X_i is a vector of individual controls including our main demographic variables (gender, age, nationality, district of residence, level of education), previous occupation, and length of unemployment. We will then enrich this vector with beliefs variables collected in the survey. As we ask beliefs after the scenarios, we will check that the scenarios received do not affect these beliefs by showing that the average beliefs are similar across the 3 blocks of scenarios.

We will estimate this regression using a:

- Simple conditional logit model: used to estimate baseline results on the average demand for training and training features.
- Mixed logit model: used to introduce taste heterogeneity between respondents.

In these estimates, we will control for an opt-out specific constant as well as a dummy for the block of scenarios.

The ratio of the estimated parameters to the parameter associated with the financial incentive of the training program measures the WTP for each feature of the training program. We could also measure the WTP as a function of the training length (WTP in terms of time investment rather than financial compensation). For example, the WTP for having employer involvement in the training would be estimated as:

$$dU = \beta_2 d(compensation) + \beta_3 d(empinvol) = 0$$
$$\Rightarrow WTP(empinvol) = \frac{\delta empinvol}{\delta compensation} = -\frac{\beta_4}{\beta_2}$$

Finally, we will exploit the richness of our data to explore the heterogeneity of WTP for training features with respect to observable characteristics of the respondents (e.g., in terms of age, gender, or length of unemployment) and their perceptions about the occupations we ask about. We will also investigate to what extent providing information on the fact that occupation ranked #2 is in shortage affects training preferences.

4.4 Do job characteristics of the target occupation matter for the decision to reskill?

In this part of our analysis, we will proceed very similarly to the part where we investigate the importance of training characteristics, but focusing on *job characteristics* instead. Specifically, we will first estimate the share of respondents who change their decision across different scenarios, indicating to what extent job features matter in general. Second, we will estimate the WTP for different job characteristics in the same way as we did it for estimating WTP for training characteristics. Third, we will look at heterogeneity based on demographics, labour market characteristics, and beliefs.

Ultimately, our aim is to compare the size of WTP measures for job characteristics with that of WTP estimates for training features.

4.5 Self-selection into employer-provided trainings

An additional research question we would like to explore in this experiment is the extent to which different groups of unemployed jobseekers self-select into employer-provided trainings. The involvement of employers in demand-driven vocational training programs tends to be highlighted as an important success factor of these programs. However, such trainings might be considered so effective specifically because participants are selected in some way (e.g., "cream-skimming"). We want to explore this question by investigating whether highlighting the fact that a training is provided in cooperation with an employer induces different types of jobseekers to enrol.

To do so, we include a small randomized information intervention at the end of the survey. In particular, when respondents are provided information about the trainings that seem to be particularly of interest to them based on their answers (see incentivization of truthful revelation described in section 2.1.2), we randomize whether the involvement of employers is mentioned in this message. In this context, the treatment group is told:

> *It seems that you are interested in finding a job in [Occupation ranked #1], and that you are potentially interested in following trainings. On this page, you can find trainings for occupation X, that seem to be of particular interest to you. Some of these trainings are organized in partnership with local employers, who commit to hiring most of the trainees at the end.*

The control group, in contrast, gets the message:

It seems that you are interested in finding a job in [Occupation ranked #1], and that you are potentially interested in following trainings. On this page, you can find trainings for occupation X, that seem to be of particular interest to you.

By looking at the profiles of individuals who click on the link with information on the trainings, we will be able to say something about the self-selection of individuals into trainings with employer involvement. In addition, we will be able to correlate the WTP for employer involvement with this "revealed choice" behavior, to further validate our experimental measures.