



An  
Phríomh-Oifig  
Staidrimh

Central  
Statistics  
Office

# Standard Report on Methods and Quality for Educational Longitudinal Database (ELD)

**Standard Report**  
**on**  
**Methods and Quality (v1)**  
**for**  
**Educational Longitudinal Database (ELD)**

This documentation applies to the reporting period:

**2019**

Last edited: 17-08-2020

CENTRAL STATISTICS OFFICE

Skehard Road, Cork

021 4535000

[www.cso.ie](http://www.cso.ie)

# Table of Contents

## 1 Overview

## 2 General Information

- 2.1 Statistical Category
- 2.2 Area of Activity
- 2.3 Organisational Unit Responsible, Persons to Contact
- 2.4 Objectives and Purpose; History
- 2.5 Periodicity
- 2.6 Client
- 2.7 Users
- 2.8 Legal basis

## 3 Statistical Concepts, Methods

- 3.1 Subject of the Statistics
- 3.2 Units of Observation/Collection Units/Units of Presentation
- 3.3 Data Sources
- 3.4 Reporting Unit/Respondents
- 3.5 Type of Survey/Process
- 3.6 Characteristics of the Sample/Process
  - 3.6.1 Population and Sampling Frame
  - 3.6.2 Sampling Design
- 3.7 Survey Technique/Data Transfer
- 3.8 Questionnaire (including explanations)
- 3.9 Participation in the Survey
- 3.10 Characteristics of the Survey/Process and its Results
- 3.11 Classifications used
- 3.12 Regional Breakdown of Results

## 4 Production of the Statistics, Data Processing, Quality Assurance

- 4.1 Data Capture
- 4.2 Coding
- 4.3 Data Editing
- 4.4 Imputation (for Non-Response or Incomplete Data Sets)
- 4.5 Grossing and Weighting
- 4.6 Computation of Outputs, Estimation Methods Used
- 4.7 Other Quality Assurance Techniques Used

## 5 Quality

- 5.1 Relevance
- 5.2 Accuracy and Reliability
  - 5.2.1. Sampling Effects, Representativity
  - 5.2.2. Non-Sampling Effects
    - 5.2.2.1 Quality of the Data Sources used
    - 5.2.2.2 Register Coverage
    - 5.2.2.3 Non-response (Unit and Item)
    - 5.2.2.4 Measurement Errors
    - 5.2.2.5 Processing Errors
    - 5.2.2.6 Model-related Effects
- 5.3 Timeliness and Punctuality
  - 5.3.1 Provisional Results
  - 5.3.2 Final Results
- 5.4 Coherence
- 5.5 Comparability
- 5.6 Accessibility and Clarity
  - 5.6.1 Assistance to Users, Special Analyses
  - 5.6.2 Revisions
  - 5.6.3 Publications
    - 5.6.3.1 Releases, Regular Publications
    - 5.6.3.2 Statistical Reports
    - 5.6.3.3 Internet
  - 5.6.4 Confidentiality

## 6 Additional documentation and publications

### 1 Overview

The Educational Longitudinal Database (ELD) is a statistical framework for the compilation and analysis of learner outcomes over many years. The ELD provides the basis for a series of projects that the CSO has established in collaboration with Irish public sector bodies to examine learner outcomes across a range of educational levels and programmes. The ELD is produced by matching datasets on learners that have completed courses

or programmes to other datasets which describe their outcomes in subsequent years. The data sources used to describe learner outcomes include employment and self-employment datasets from the Revenue Commissioners, benefits data from the Department of Social Protection, and data on educational participation from the Department of Education and a number of state agencies, including the Higher Education Authority (HEA), Quality and Qualifications Ireland (QQI) and SOLAS.

## **2 General Information**

### **2.1 Statistical Category**

This project is based on administrative data sources.

### **2.2 Area of Activity**

Learner outcomes and progression through the education system.

### **2.3 Organisational Unit Responsible, Persons to Contact**

Statistical Systems Coordination Unit (SSCU)

Brendan O'Dowd 01-4984057

Kieran Culhane

[sscu@csso.ie](mailto:sscu@csso.ie)

### **2.4 Objectives and Purpose; History**

The ELD provides a framework for analysing the outcomes of learners following educational courses, as well as the movements of learners through the educational system. Publications such as 'Higher Education Outcomes' and 'Further Education Outcomes' use the ELD as their basis. The purpose of the ELD is to produce useful information for educational organisations (schools, universities, regulatory organisations, etc.), students (& their families), guidance counsellors, employers and policy makers. The work also demonstrates the statistical potential of administrative data, and makes the case for greater integration and alignment of administrative datasets, as recommended under the National Data Infrastructure (NDI). The ELD can also be the basis for collaborative projects between the CSO and other organisations. One collaboration between the CSO and the HEA involved a researcher from the HEA using a customised portion of the ELD to produce an in-depth report on graduate earnings: 'Labour Market Earnings for Graduates in their Early Careers' (December 2019).

### **2.5 Periodicity**

Publications such as Higher Education Outcomes and Further Education Outcomes will be produced on a biannual basis. Other collaborative projects may be produced on an ad hoc basis.

### **2.6 Client**

General National Requirement. Some projects are carried out in collaboration with other educational organisations, including the HEA and Solas.

### **2.7 Users**

Learners

Policy makers

Educational organisations

Guidance counsellors

Employers

### **2.8 Legal basis**

Section 11 of the Statistics Act, 1993

## **3 Statistical Concepts, Methods**

### **3.1 Subject of the Statistics**

The ELD provides a framework for analysing learner outcomes and progression of learners through the education system. It uses a range of administrative data sources covering education, employment and benefits, and establishes the principal activities of individuals within each calendar year. Activity categories include 'substantial employment' and 'enrolment in education', with further granularity available such as the industry sectors of employment and the fields of study that learners are engaged in. Earnings of learners/graduates who are in employment can also be analysed.

### **3.2 Units of Observation/Collection Units/Units of Presentation**

The unit of observation is a learner or graduate in the Irish educational system, as defined by records from the HEA, QQI and the Department of Education & Skills. The contents of the ELD are adjusted according to the cohort required for each individual study.

### **3.3 Data Sources**

All data sources are Analysis Tier (pseudonymised) versions of administrative datasets.

- Business Register Data
- Person Income Register (PIR)

- Central Record System (CRS) Client details, from DEASP
- Department of Education teaching and other staff information
- Solas client and course details
- HEA student and course details, including enrolment and graduation records, as well as Springboard records.
- QQI award data
- Income Tax Form 11 (Self-employment data) from Revenue
- Jobseekers Longitudinal Dataset (JLD)
- Real Time Revenue Data (PMOD)
- Pobal Programmes Implementation Platform (PIP) – Childcare providers
- Post Primary Pupils Database from the Department of Education
- Leaving Certificate results from the SEC
- Employee level P35 data, including linked variables from CSO and DEASP
- SUSI grant application and payment data
- Teaching Council register of teachers

### **3.4 Reporting Unit/Respondents**

Respondents are as follows:

All datasets are pseudonymised ‘analysis tier’ datasets provided and managed by the CSO’s Administrative Data Centre (ADC). The primary ‘source tier’ files come from the following organisations:

- Department of Education and Skills
- HEA (Higher Education Authority)
- QQI (Quality & Qualifications Ireland)
- Solas
- Pobal
- SEC (State Examinations Commission)
- Revenue Commissioners
- Department of Employment Affairs and Social Protection
- The Teaching Council
- SUSI

### **3.5 Type of Survey/Process**

The ELD is based on administrative data sources.

### **3.6 Characteristics of the Sample/Process**

#### **3.6.1 Population and Sampling Frame**

The ELD is based on a full population of learners for each analytical cohort (less records with missing or invalid PPSN).

#### **3.6.2 Sampling Design**

Not applicable

### **3.7 Survey Technique/Data Transfer**

Data is transferred by the various organisations to the CSO’s Administrative Data Centre (ADC) using a Secure File Transfer Protocol (SFTP).

### **3.8 Questionnaire (including explanations)**

Not applicable

### **3.9 Participation in the Survey**

Not applicable

### **3.10 Characteristics of the Survey/Process and its Results**

The ELD integrates data from a wide range of sources. Much of the processing involves harmonisation of these data sources and the extraction of categorical activities for each person in each calendar year.

From educational data sources (HEA, Solas, QQI, post-primary pupils database, etc.), we identify the time periods in which individuals were engaged in education, when qualifications were received, and the level and field of study of the courses undertaken.

From employment data sources (P35, PMOD, PIR, IT form 11, JLD), we examine whether individuals meet our criteria for ‘substantial employment’ in each calendar year. We also identify weekly earnings from a main employer.

From welfare data sources (CRS, JLD) we examine other activities within the state, determine whether individuals had ‘substantial unemployment’ within the year. This data is also used for relational information and as a component of the definition for ‘substantial employment’.

The Pobal data source is used to identify employers which are approved by this body.

The Teaching Council data sources is used to analyse outcomes for learners who go on to become teachers.

The business register is used to add additional information to employment activities.

Most publications include some of the following outcome categories:

- Substantial employment – meeting threshold levels of weeks in employment and pay, or income in the case of self-employment
- Enrolled in education – levels of education included here depends on the nature of the report, and these are drawn from a number of administrative datasets from different educational organisations.
- Neither employment nor education – people who are not in either of the previous two categories but who appear somewhere in the administrative data for that year; e.g. they have a record of receiving some benefit, or they worked for a period of time below the threshold required for “substantial” employment.
- Not captured – people who do not appear anywhere in the administrative data for the year in question. It is possible that these people have emigrated, but there is no administrative indicator to confirm emigration.

### 3.11 Classifications used

Below is an example of provider-defined codes.

- Sex
- Field of study (ISCED)
- NFQ level
- Institution and institution type
- Geographic region (NUTS)
- NACE

ELD-defined codes:

- Young/Mature graduates – based on age and qualification type
- Outcome activity – based in particular on employment and education records, but including input from all administrative sources
- Substantial employment, substantial self-employment, substantial unemployment – based on income and duration
- ‘Main employer’ – based on individuals’ employment history across all employment records within calendar year
- Weekly earnings – based on income from Revenue data
- Churn – number of employers within a given period.

### 3.12 Regional Breakdown of Results

Some data is broken down by NUTS 3 region in Ireland.

<b>NUTS 3:</b>					
<b>Border</b>	Cavan	<b>Mid-West</b>	Clare	<b>Dublin</b>	Dublin City
	Donegal		Limerick		Dun Laoghaire-Rathdown
	Leitrim		Tipperary		Fingal
	Monaghan				South Dublin
	Sligo	<b>South-East</b>	Carlow		
			Kilkenny	<b>Mid-East</b>	Kildare
<b>Midland</b>	Laoighis		Waterford		Meath
	Longford		Wexford		Wicklow
	Offaly				Louth
	Westmeath	<b>West</b>	Galway		
			Mayo	<b>South-West</b>	Cork
			Roscommon		Kerry

## 4 Production of the Statistics, Data Processing, Quality Assurance

### 4.1 Data Capture

Datasets are transferred by a number of organisations to the Administrative Data Centre (ADC) within the CSO, which carries out initial processing of the data, including the preparation of pseudonymised versions of the datasets. These do not include personal identifiers, but may be linked via a protected identifier key (PIK).

### 4.2 Coding

The ELD includes a wide range of classifications, some of which are defined by the data provider and others which are defined through the framework of the ELD (see section 3.11). Classifications produced by data providers are rarely edited, but are subject to quality checks and corrections. All of these are managed in SAS, and classifications for Statbank files are also managed in CARS.

### **4.3 Data Editing**

All input datasets are edited in SAS to ensure compatibility and to enable joining. Quality checks such as frequency counts are carried out to identify potential problems. Significant outliers or unexplained trends or spikes are brought to the attention of data providers for clarification. Administrative datasets are joined for each calendar year for a set of CSOPPSN which is defined for each ELD project.

### **4.4 Imputation (for Non-Response or Incomplete Data Sets)**

Imputation is not used for ELD publications produced by the CSO due to the high level of PPSN coverage within the administrative sources. However, rates of missing PPSN are given for a range of categorical variables to give users a sense of how coverage varies.

### **4.5 Grossing and Weighting**

N/A

### **4.6 Computation of Outputs, Estimation Methods Used**

All outputs are produced in SAS (a statistical software package).

### **4.7 Other Quality Assurance Techniques Used**

Quality checks are conducted on all input datasets, including frequency counts, detection of outliers and inspection of trends. Issues are examined usually with the aid of the data provider.

## **5 Quality**

### **5.1 Relevance**

The ELD is a comprehensive source of information on the performance of Ireland's education sector and the outcomes for learners within it. It is a powerful tool for examining the interconnected nature of the various parts of the education system. It can be used to examine equity of access and progress through the system. It can be used to benchmark performance of different institutions and organisations. It provides a basis for examining graduate supply, and which sectors of the economy these graduates are finding employment. It can be used to examine and compare the benefits of pursuing different courses and programmes that are available to learners.

The Education Statistics Liaison Group (ESLG) brings together the main educational organisations in Ireland, including the CSO, the Department of Education & Skills, the HEA, Solas, the SEC, QQI, the Department of Public Expenditure and Reform and the Teaching Council. The projects and future direction of the ELD is discussed at the ESLG.

### **5.2 Accuracy and Reliability**

#### **5.2.1. Sampling Effect & representivity**

Not applicable

#### **5.2.2. Non-Sampling Effects**

*Note: Give descriptions, measurements and explanations under each of the following headings*

##### **5.2.2.1 Quality of the Data Sources used (other than survey register)**

The principal quality issue relating to the administrative datasets used in the ELD is missing/invalid PPSN. It is usually not possible to rectify this issue. The proportion of missing or invalid PPSN is given for a range of analysis variables. Some datasets have inconsistencies in the categorisation of certain variables. This may arise due to a change in classification structure (as happened to the ISCED field of study classification in 2012), or because classifications are applied by a number of different bodies using their own interpretations. Quality checks such as frequency counts are carried out to identify classification issues. Significant outliers or unexplained trends or spikes are brought to the attention of data providers for clarification.

##### **5.2.2.2 Register Coverage**

Not applicable

##### **5.2.2.3 Non-response (Unit and Item)**

A proportion of learner records do not have an associated PPSN, and these records cannot therefore be linked to other administrative sources. The distribution of missing or invalid PPSN is described within each publication.

##### **5.2.2.4 Measurement Errors**

Some administrative datasets have a proportion of records with missing or invalid PPSN, which cannot be linked to other datasets. Some datasets include categorisations which can be open to interpretation, e.g. field of study variables in HEA data. These are managed by carrying out quality checks on the data, examining distributions and trends, and discussing any anomalies with data providers.

#### **5.2.2.5 Processing Errors**

The scope for processing errors is limited, in that the data is not processed manually. Inconsistencies in coding and unusual trends in both input and output data are discussed with experts and data providers. Results are checked for consistency across different categorical breakdowns.

#### **5.2.2.6 Model-related Effects**

N/A

### **5.3 Timeliness and Punctuality**

#### **5.3.1 Provisional Results**

It is not envisaged that publications with provisional results will be produced. However, data relating to a certain calendar year may be slightly adjusted in newer publications due to the arrival of new administrative data relating to that year. In general this is rare as new administrative data usually relates to newer years, but it is possible that records relating to older years will arrive, particularly with self-employment data (IT form 11).

#### **5.3.2 Final Results**

The Higher Education Outcomes and Further Education Outcomes publications will be produced on a biannual basis. The most recent release dates for these publications is 18 July 2019. Other publications will be produced on an as-required basis, such as the collaborative report 'Labour Market Earnings for Graduates in their Early Careers' (Dec 2019).

### **5.4 Coherence**

The only other comparable source of data on this subject is the series of publications by the HEA called 'What do Graduates do?'. This HEA series is based on survey data taken at a single point in time each year, and is therefore different from the ELD in a number of ways. However, these differences are scrutinised and discussed with the HEA on an ongoing basis.

### **5.5 Comparability**

All data should be consistent from year to year. However, due to changes in the availability or composition of the administrative datasets used, there may be slight variations in some outcomes. Issues such as these will be highlighted in the publications.

The analysis of learner outcomes using administrative data is not yet widespread, and the types of analysis that are possible depends on the nature of the administrative data available, which varies from one country to another. However, we regularly review outputs from other statistical agencies that re producing reports on this subject, and where possible make reasonable efforts to align our definitions and categorisations.

### **5.6 Accessibility and Clarity**

#### **5.6.1 Assistance to Users, Special Analyses**

Detailed background notes are included with each publication and a methodological document is also provided for the ELD. Support and clarification are also provided by phone and email. Publications are not translated but infographics are.

#### **5.6.2 Revisions**

Published data may be subject to very minor revisions, for example due to receipt of updated self-employment data which may relate to periods up to four years prior. In such instances a note will be included with Statbank tables and background information to the publications.

#### **5.6.3 Publications**

##### **5.6.3.1 Releases, Regular Publications**

Higher Education Outcomes  
Further Education Outcomes  
Post Primary Outcomes

##### **5.6.3.2 Statistical Reports**

Labour Market Earnings for Graduates in their Early Careers (HEA collaboration)

##### **5.6.3.3 Internet**

Some publications are available here: <https://www.cso.ie/en/statistics/education/>

Report arising from collaboration with the HEA available here: <https://hea.ie/2019/12/11/labour-market-earnings-for-graduates-in-their-early-careers/>



Higher Education Outcomes Statbank files:

[https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Higher%20Education%20Outcomes/Higher%20Education%20Outcomes\\_statbank.asp?sp=Higher%20Education%20Outcomes&Planguage=0](https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Higher%20Education%20Outcomes/Higher%20Education%20Outcomes_statbank.asp?sp=Higher%20Education%20Outcomes&Planguage=0)

Further Education Outcomes Statbank files:

[https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Further%20Education%20Outcomes/Further%20Education%20Outcomes\\_statbank.asp?sp=Further%20Education%20Outcomes&Planguage=0](https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Further%20Education%20Outcomes/Further%20Education%20Outcomes_statbank.asp?sp=Further%20Education%20Outcomes&Planguage=0)

Post Primary Outcomes Statbank files:

[https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Post%20Primary%20Outcomes/Post%20Primary%20Outcomes\\_statbank.asp?sp=Post%20Primary%20Outcomes&Planguage=0](https://statbank.cso.ie/px/pxeirestat/Database/Eirestat/Post%20Primary%20Outcomes/Post%20Primary%20Outcomes_statbank.asp?sp=Post%20Primary%20Outcomes&Planguage=0)

#### **5.6.4 Confidentiality**

Only analysis tier (pseudonymised) datasets are used in the analysis.

All counts are rounded to the nearest ten.

Only median and quartile figures are given for earnings, and these are rounded to the nearest €5.

## **6 Additional documentation and publications**

Below are some related publications by other organisations.

What do Graduates Do? – A series of publications by the HEA based on survey data

A number of publications by Stats NZ (the official data agency of New Zealand) use administrative data to examine learner outcomes. The database used is known as the Integrated Data Infrastructure or IDI. Some publications can be found at these links:

<https://www.educationcounts.govt.nz/publications/>

<https://cdm20045.contentdm.oclc.org/digital/collection/p20045coll17>