# An Evaluation of the Look-Up Approach to Occupation Coding

Helena Koerber, Matt Brown, Lisa Calderwood

UCL
CENTRE FOR LONGITUDINAL STUDIES
SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION
UKRI Economic and Social Research Council

# Overview:

- Background
- Coding approaches
- Data: survey experiment
- Key findings

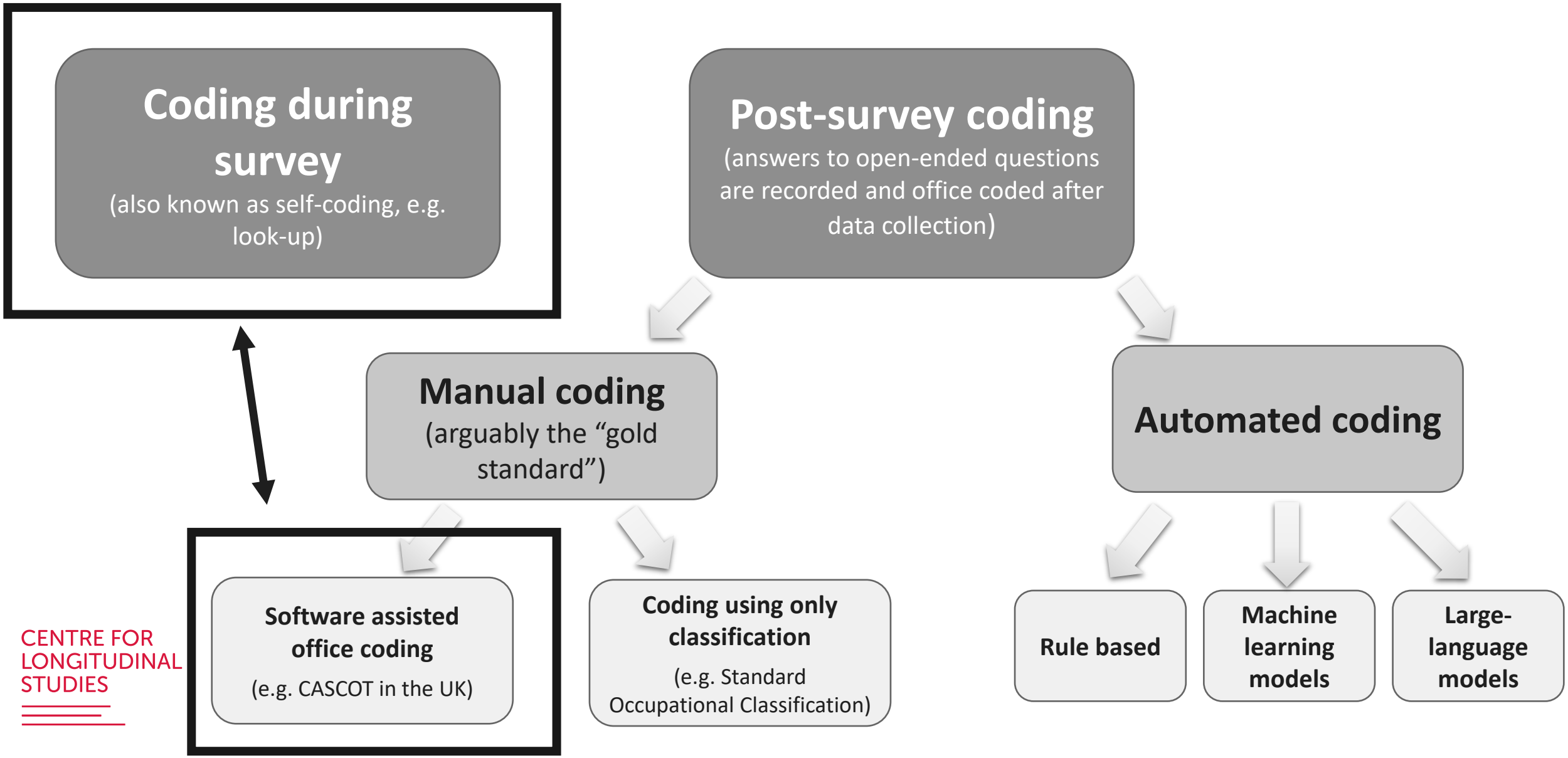
# Background

- Occupation is a core measurement in social surveys:
  - Indicator of socio-economic status
  - Strongly linked to income, health, and lifestyle
- Traditional Measurement Approach:
  1. **Interviewers** asking **open questions** to collect job title and a description of duties (Lyberg & Dean, 1992)
  2. **Interviewers** ensure the necessary information is provided (Conrad et al., 2016)
  3. **Software assisted** manual coding by specialist **office-based coders** to a **standard classification** (e.g., SOC 2020)

# Background

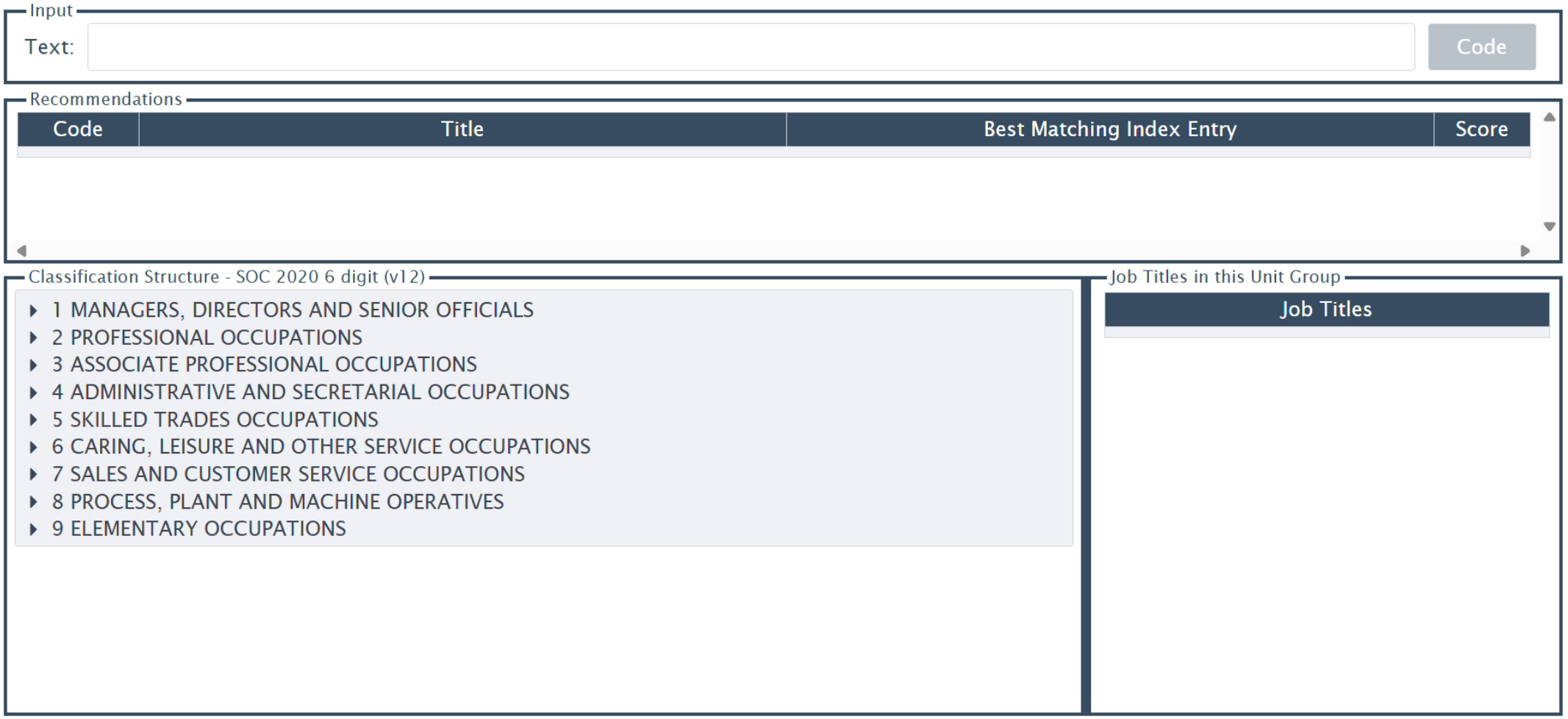
- Methodological challenges:
  - Job titles/ descriptions are highly diverse
  - Same job may be described in multiple ways
  - In **self-completion surveys**, the absence of interviewers can have a negative impact on the quality of the collected data for coding (Conrad et al., 2016)
- Ongoing research aims to improve validity and reliability of occupation coding in surveys

# Focus of this research:



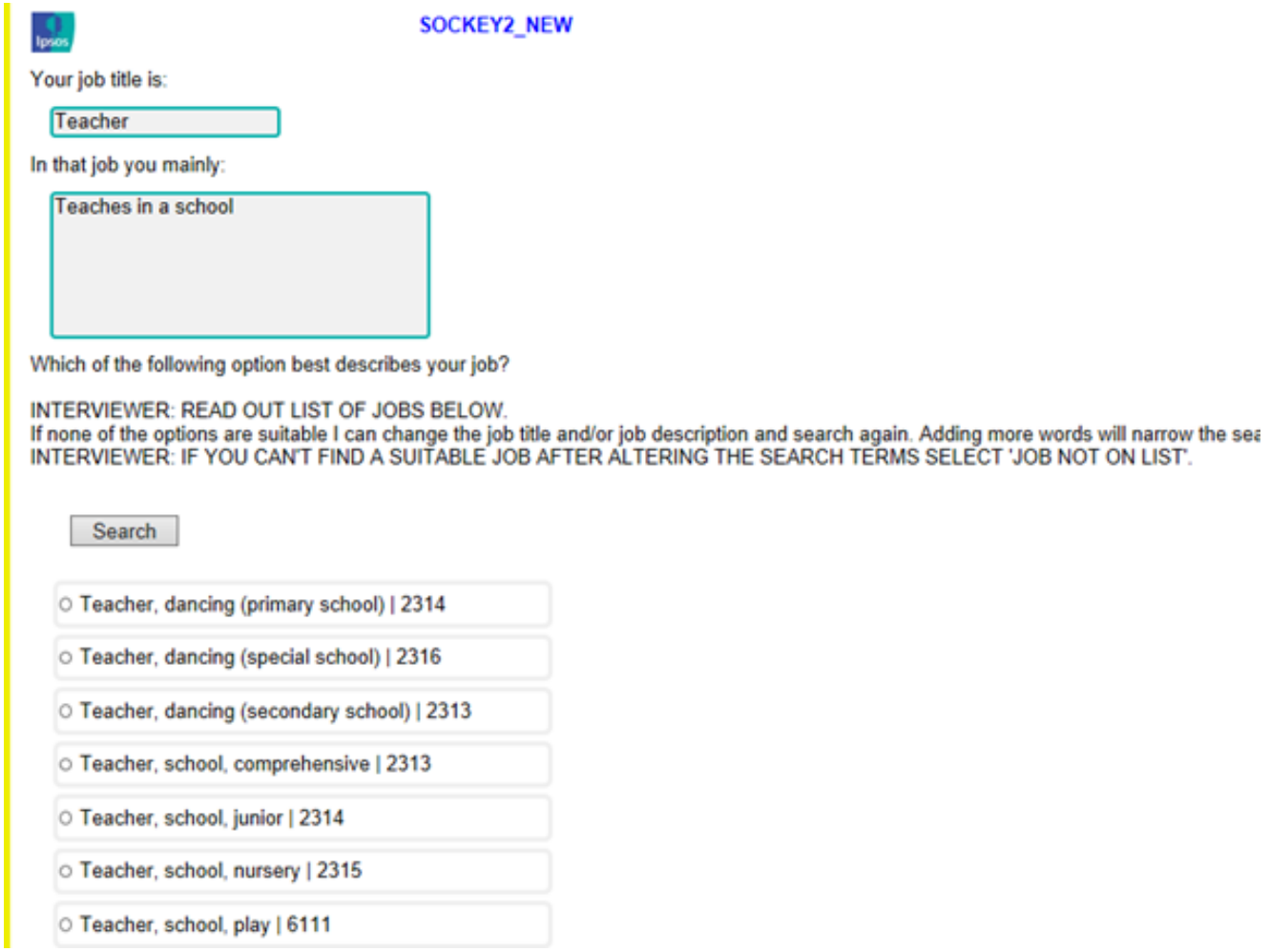

CENTRE FOR LONGITUDINAL STUDIES

# CASCOT assisted office coding




CENTRE FOR LONGITUDINAL STUDIES

# The Look-Up Approach




CENTRE FOR LONGITUDINAL STUDIES

# Coding Approaches:

**Office Coding:**

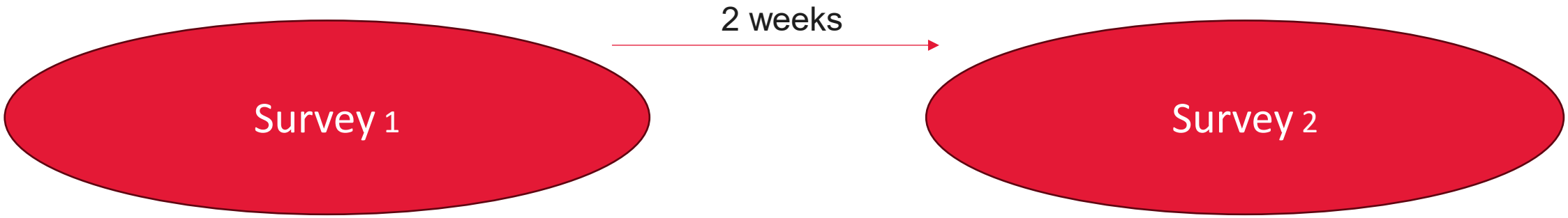
Reliable but time consuming and costly

**Look-up tool:**

New, fast and cost efficient

But reliable?

# Study Design: Mode experiment




- Random mode allocation: online, video, in-person
- Look-up coded + 2x office coded: coded to 4-digits
- Question to determine whether participants had changed jobs since survey 1
- Exclusion criteria: changed jobs or not working at time of either survey
- Fieldwork conducted by IPSOS
- Non-probability sample of 20-to 40-year-olds

CENTRE FOR LONGITUDINAL STUDIES

# Mode Groups:

| Mode Group | Frequency |
|---|---|
| **Online/ Online** | 135 |
| **Online/ Video** | 136 |
| **Online/ In Person** | 130 |
| **Video/ Online** | 134 |
| **Video/ Video** | 137 |
| **Video/ In Person** | 139 |
| **In Person/ Online** | 122 |
| **In Person/ Teams** | 133 |
| **In Person/ In Person** | 129 |
| **Total** | 1195 |

CENTRE FOR LONGITUDINAL STUDIES

# Measuring Occupation

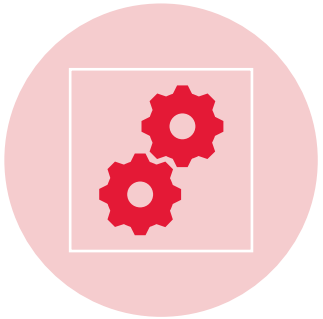
Look-Up

*"What is your job title?"*
*"Please tell us keywords which describe what you do in your job"*
*"How well does the selected option describe your current job?"*

Open-Text Job Description

*"This approach to collecting information about your job is new and we are testing it out. To help us check whether it is working, could you also describe in your own words what you mainly do in your job? Please describe in detail (for example the type of work, the department you are in, and what level you work at)."*

# Evaluation of Look-Up Tool: Quality Indicators:



**Coding Success:** Successful selection of a job code. Office coding as benchmark.

**Subjective Accuracy:** Participant's perceived accuracy of how well the selected job code reflects their actual job.

**Agreement with Office Coding:** The degree of agreement between the participant-selected job code (via the Look-up Tool) and the job code assigned by an expert office coder.

**Coding Consistency:** The consistency of participant-selected job codes using the Look-up Tool across two experimental waves. Office coding as a benchmark.

CENTRE FOR LONGITUDINAL STUDIES

# Research Questions

- **RQ1:** How successful are respondents at **selecting an occupation code** using the look-up tool and how does this vary by mode?
- **RQ2:** How do respondents rate the **accuracy of the occupation code** they select using the look-up tool and how does this subjective accuracy vary by mode?
- **RQ3:** How closely do respondent selected look-up codes align with codes assigned by **office coders** and how does this vary by mode?
- **RQ4:** Among respondents who have not changed jobs, how **consistent** are respondents at selecting look-up codes across two time points and how does this compare with consistency of office coding? How does this consistency vary based on whether participants **switched modes**?

# Results: coding rates by mode wave 1

| Mode | Look-up coding rate | Office coder 1 coding rate |
|---|---|---|
| **Online** | 86.3% | 100% |
| **In-person** | 88.1% | 100% |
| **Video** | 86% | 100% |
| **Total** | **86.6%** | **100%** |

CENTRE FOR LONGITUDINAL STUDIES

# Results: self-reported accuracy wave 1

| Mode | Subjective Accuracy | | |
|---|---|---|---|
| | Very Well | Fairly well | Not very/ not at all well |
| **Online** | 45.6% | 51.1% | 3.4% |
| **In-person** | 54.5% | 42.5% | 2% |
| **Video** | 64.1% | 34% | 2% |
| **Total** | **54.8%** | **42.5%** | **2.8%** |

CENTRE FOR LONGITUDINAL STUDIES

# Results: agreement with office coding

| Mode | Agreement the look-up & 1st office coder | | | | Agreement 1st office coder & 2nd office coder | | | |
|---|---|---|---|---|---|---|---|---|
| | 1-digit | 2-digit | 3-digit | 4-digit | 1-digit | 2-digit | 3-digit | 4-digit |
| **Online** | 80.1% | 74.8% | 72% | 65.7% | 98.3% | 96.8% | 96.5% | 95% |
| **In-person** | 75% | 70.8% | 68.8% | 63.4% | 98.4% | 97.2% | 96.9% | 95.1% |
| **Video** | 77.9% | 74.2% | 69.7% | 65.1% | 98.4% | 97.3% | 97.8% | 95.3% |
| **Total** | **77.7%** | **73.3%** | **70.2%** | **64.7%** | **98.4%** | **97.1%** | **96.8%** | **95.1%** |

CENTRE FOR LONGITUDINAL STUDIES

# Results: coding consistency across two waves

| Digit Level | Look-up Coding consistency | Office Coding consistency |
|---|---|---|
| **1-digit** | 76.1% | 83.1% |
| **2-digit** | 70.7% | 80% |
| **3-digit** | 67.1% | 78.2% |
| **4-digit** | 60% | 73.7% |

CENTRE FOR LONGITUDINAL STUDIES

# Results: coding consistency: mode switching

## Look-Up

| Digit | Coding consistency **same mode** | Coding consistency **switch video & in-person** | Coding consistency: **switch online interviewer-assisted** | Rao-Scott corrected Chi2 | p-value |
|---|---|---|---|---|---|
| 1-digit | 76.7% | 79.9% | 73.7% | .9 | .38 |
| 2-digit | 73.7% | 74.4% | 66.3% | 1.9 | .14 |
| 3-digit | 71.9% | 69.7% | 62% | 2.7 | .07 |
| 4-digit | 66.3% | 58.1% | 55.9% | 2.6 | .07 |

## Office Coding

| Digit | Coding consistency **same mode** | Coding consistency: **switch video & in-person** | Coding consistency: **switch online interviewer-assisted** | Rao-Scott corrected Chi2 | p-value |
|---|---|---|---|---|---|
| 1-digit | 88.4% | 82.9% | 79.2% | 4 | .02 |
| 2-digit | 85.9% | 82.1% | 74.4% | 5.7 | <.01 |
| 3-digit | 84.4% | 80.3% | 72.6% | 5.9 | <.01 |
| 4-digit | 77.8% | 75.8% | 69.6% | 2.6 | .08 |

CENTRE FOR LONGITUDINAL STUDIES

# Error Sources:  Office Coding

Kim & Kim (2025) identified two sources of error that can lead to low reliability:

1. low intercoder reliability, where coders assign different codes to the same description
2. Alterations in respondent descriptions

CENTRE FOR LONGITUDINAL STUDIES

# Error Sources:  Look-up

1. Different input into look-up tool leads to different output: participants choose different occupation that resembles their occupation the closest
2. Same input, same list generated by the look-up tool, but participants select different job from the list

Participants may not fully understand how occupation codes are structured or what they represent (Watson & Jones, 2025).

# Key findings and conclusion

- Look-up approach has potential to reduce the need for office coding – but results suggest it could not replace manual coding
- Low levels of agreement with office coding and low levels of consistency raise concerns with the potential quality
- We are currently looking into causes of low coding consistency
- Further work could develop a more user-friendly code frame which uses terms people use to describe their jobs
- AI approaches are also being explored for real-time coding

# Thank you for your attention!

**Helena Koerber, Matt Brown, Lisa Calderwood**

CENTRE FOR LONGITUDINAL STUDIES

SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Economic and Social Research Council