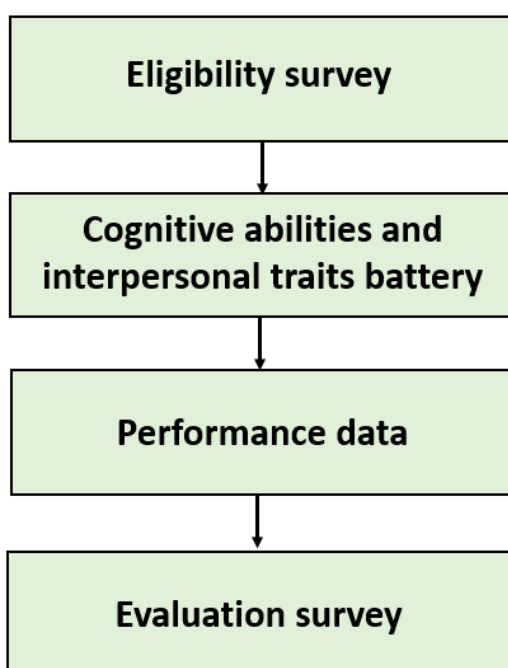


## SMART PROJECT DATA COLLECTION METHODS AND CONTEXT DOCUMENTATION

The SMART project was designed to collect and triangulate different types of data to be analysed via quantitative and qualitative approaches. The experiment set up was organised in different sequential stages, which replicate some components of previous experiments, mostly the SMART pilot study (reported in [Davitti and Sandrelli 2020](#)). This document outlines the sequential steps of data collection (4 steps) during the SMART study and the corresponding data folders deposited with open access.



**Step 1: Eligibility Survey** was used to collect relevant information about each participant before the start of the upskilling course.

**Step 2: Battery of Cognitive Tests and Interpersonal Trait Measures:** Participants were tested on cognitive abilities and interpersonal traits before undertaking specific training in interlingual respeaking. This enabled the researchers to correlate participants' interlingual respeaking performance with their existing cognitive attributes and interpersonal traits. Additional cognitive measures were taken after the training course on reading span, switching skills, and sustained attention to see any possible cognitive changes based on the training course.

**Step 3: Performance data:** After completing the upskilling course to gain specific procedural skills in interlingual respeaking, participants were tested in this practice with different types of content (i.e., characterised by speed, planned/unplanned delivery, and multiple speakers).

**Step 4: Evaluation Survey:** This was used to collect relevant information about participants' overall satisfaction with their performance and the training received. Participants who completed all stages received personalised feedback on their test performance and a certificate that allowed participation to count as Continuous Professional Development (CPD).

The experiment design was developed by the Principal Investigator (Davitti) and Co-Investigators (Evans, Sandrelli, Romero-Fresco) on the project, with input from the Advisory Board (industry and academic) and Research Fellows on the project. For further information on the team, visit the [relevant page](#) (Team) on the dedicated [SMART website](#). For any **queries or questions**, do contact: Dr Elena Davitti, Associate Professor in Translation Studies at the Centre for Translation Studies, University of Surrey  
**Email:** [e.davitti@surrey.ac.uk](mailto:e.davitti@surrey.ac.uk)

Below, you can find further information on the type of data collected during each of the four steps.

### Eligibility Survey

In step 1, participants were expected to complete an online survey structured into three main parts: I. informed consent questions to participate in the study; II. eligibility questions to determine if exclusion criteria apply (namely questions on the availability to join the course, language eligibility, professional experience, and technical equipment specifications); III. If applicants met these eligibility criteria, they were asked further questions, including demographic information, language proficiency, qualifications, experience in relevant fields, skills and training profile for thorough background profiling, as well as their expectations and reasons for wanting to join the study. The complete survey lasted approximately 20-30 minutes. The rationale for administering the survey before accessing the training course was to ensure potential participants met all eligibility criteria, particularly their status as language professionals. Demographic data was only collected from those participants who met all the other inclusion criteria. The Eligibility Survey was administered in Qualtrics.

### Battery of Cognitive Tests and Interpersonal Trait Measures

Step 2 required the participants to complete a battery of psychological tests comprising standard cognitive tasks and questionnaire-based trait measures, as detailed in Table 1. The tasks were administered in the order shown. This was arranged so that attentionally demanding tasks alternated with less demanding measures to ease pressure on participants. These were fundamental to acquiring information about participants' cognitive resources and individual differences. This component represented one of the elements of the originality of SMART with respect to previous studies.

The whole test battery took approximately 60 minutes to complete online. The participants were given a link to the study that automatically guided them through all the questionnaires (created in Qualtrics) and cognitive tasks that were created in Pavlovia. All tasks were seamlessly integrated to ensure a smooth flow, automatically guiding participants from one task to the next. Before starting the battery, each participant had a quick introduction to the experiment over Zoom with a Cognitive Psychologist. As part of this call, the participants were also tested on their verbal fluency skills (further details can be seen in Table 1).

Completing the test battery was a pre-requisite to be given access to training course material. The participants were tested before the interlingual respeaking training course on all the cognitive measures and subjective measures (i.e., self-reported questionnaires) mentioned in Table 1. After the training, we also measured the participants' reading span, digit span, and sustained attention to see how the training affected these cognitive functions by comparing their pretesting cognitive performance with the performance after the training on interlingual respeaking. At the end of this document, we added the scoring system for the cognitive tasks and questionnaires on interpersonal skills in Appendix I and II.

**Table 1.** Battery of Cognitive Tests and Interpersonal Trait Measures

	Task Name	Platform	Short Description of Measure
0	Verbal Fluency Task	Zoom	Participants are given 1 min to produce as many unique words as possible within a semantic category (category fluency, e.g., animals) or starting with a given letter (letter fluency, e.g., letter F). The participant's score in each task is the number of unique correct words.
1	State-Trait Anxiety Inventory (STAI) (only trait version was used)	Qualtrics	The State-Trait Anxiety Inventory (STAI) is a commonly used measure of trait and state anxiety (Bieling et al., 1998). There are 20 questions in total (trait version), and it uses a 4-point scale (1 = not at all). Reverse scoring for anxiety-absent items and a total score of 20-80.
2	Plus-Minus Task	Qualtrics	<p>Plus-Minus Task measures switching skills by using simple mathematical equations. The participants start with adding, moving into subtraction, and finishing with a task where they alternate between additions and subtractions. A switching cost is calculated to see how well the participants alternate between two different types of calculations (i.e., additions and subtractions).</p> <p>A Switching cost = switch cost RT (time to complete switch condition) – non-switch cost RT (an average of addition and subtraction conditions).</p>
3	Brief Resilience Scale (BRS)	Qualtrics	BRS assess the ability to bounce back or recover from stress. There are six questions (3 reverse scored) on a scale of 5 items.
4	Digit Span Task (DST)	Pavlovía	<p>DST is a simple verbal working memory measure. Working memory temporarily stores and maintains information that is required for the successful completion of cognitive tasks. Participant does forward span (attention) and backward span (memory) and then must recall in ascending order.</p> <p>DST was tested both before and after the training course.</p>
5	Barratt Impulsiveness Scale (BIS)	Qualtrics	BIS measures impulsive or non-impulsive behaviours and preferences. 15 questions represent 3 subdivisions: A – Attention impulsivity, M – Motor impulsivity, and NP – Non-planning impulsivity. Seven items are reverse scored (non-impulsive statements). The scores are from 1 = rarely - 4.
6	Cognitive Flexibility Scale (CFS)	Qualtrics	CFS measures the cognitive flexibility of the person. Participants are asked to answer 12, 6-point Likert-type questions. The values of the 12 items are summed up.
7	N-back Task	Pavlovía	A widely used measure for assessing working memory function is the <i>n</i> -back task. Participants are typically instructed to monitor a series of stimuli and to respond whenever a stimulus is presented that is the same as the one presented <i>n</i> trials previously.
8	Personal Innovativeness in IT Scale (PIIT), Lopez-Bonilla and Lopez-Bonilla (2012)	Qualtrics	PIIT measures a participant's willingness to engage with Information Technology. 4 questions are scored from 1-7.

	Break		Break
9	Ten Item Personality Indicators (TIPI)	Qualtrics	The TIPI was developed as a brief inventory for the Big Five personality traits: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to New Experiences whereby participants rate 10 pairs of descriptive items such as “Disorganized, Careless” on a 7-point Likert scale.
10	The Sustained Attention to Response Task (SART)	Pavlovica	SART is a go/no-go task that requires participants to withhold behavioural response to a single, infrequent target presented amongst a background of frequent non-targets.  SART was tested both before and after the training.
11	The Work Extrinsic and Intrinsic Motivation Scale (WEIMS)	Qualtrics	WEIMS is a measure of work motivation. There are 6 subscales: Intrinsic motivation, Integrated regulation (work is part of you), Identified regulation (the individual identifies value), Introjected regulation (self-estimated guilt), External regulation (based on award) and Amotivation (lowest). A work self-determination index is also calculated (relative level of self-determination).
12	Reading Span Task (RST)	Pavlovica	RST is a complex verbal working memory test. It combines a processing component (judging the correctness of a sentence) and a storage component (memorising a series of words for a recall). Only the storage component is reported. The processing component ensures that the participant pays attention to the required task.  RST was tested both before and after the training course.
13	Five Facet Mindfulness Questionnaire (FFMQ), short version (Bohlmeijer et al., (2011)	Qualtrics	FFMQ measures five facets of mindfulness: Non-reacting (detaching), Observe, Act aware (acting mindfully), Describe, and Non-judge.

## Performance Data

The interlingual respeaking training component provided participants with approximately 25 hours of upskilling in interlingual respeaking from experts in the field (Step 3). The training was delivered as a short, self-taught online course designed as a CPD opportunity for participants and was free of charge. It was referred to as “Upskilling for Professionals: Advanced Introduction to Interlingual Respeaking” (or simply upskilling course) in documents going out to participants as this was easier to understand.

During the training, the complex technique of interlingual respeaking was broken down into core procedural skills that could be taught and trained progressively via a blending and scaffolding approach. Participants, who were all language professionals and whose eligibility had been checked during Step 1 (Eligibility survey) already possessed some of the core skills required; depending on their professional background, they needed to adjust, unlearn or acquire new ones to be able to perform interlingual respeaking.

The course aimed to identify core components of this practice, track and understand how they could be transferred and/or acquired by professionals who brought their own unique set of skills, cognitive abilities, and interpersonal traits. The final goal was to find a way to customise the future upskilling course based on each individual's background and existing skill set. This layering approach differentiated itself from previous training approaches to interlingual respeaking, in that technical skills and translational skills were gradually and incrementally built into each stage of training.

The training course comprised four blocks of learning. Each block included three components: one related to acquiring skills for software management of Dragon Professional Individual v15, one related to skills to acquire intralingual respeaking abilities and one related to skills for interlingual respeaking. To ensure the rigour of the design, some conditions were imposed on the course material so that participants could progress in a systematic manner through each task and exercise, and in the same order. Completion of each block was a prerequisite to starting the next. Upon completion of the course, participants applied the skills acquired to interlingual respeaking tests, each posing a different challenge: (1) speed, (2) planned/unplanned delivery, and (3) multiple speakers.

The course was organised and delivered via the online platform Moodle, which had been integrated with screencast technology to collect data in relation to the final testing. The data provided in the folder includes the transcript of each source speech provided and the target output, i.e., the subtitles produced by participants via interlingual respeaking. These are aligned by idea units. Analysis in SMART was conducted via a purpose-made grid to apply the NTR model (Romero-Fresco and Pöschhacker, 2017). For information about the grid and analytical approach adopted in SMART, see Davitti and Sandrelli (2020) and forthcoming SMART-related papers.

### Evaluation Survey

After completing the upskilling course and testing described above, participants completed the Evaluation Survey (approximately 20 minutes), allowing them to reflect on the training they had just taken part in (Step 4). For instance, we asked how much they felt their initial expectations and motivation had been met or fulfilled. We also asked them for feedback on the course and how it met their expectations. Finally, we also asked the participants how they hoped to use the skills that they had gained. The Evaluation Survey was administered in Qualtrics.

## APPENDIX I: Scoring of the Cognitive Tasks

### Verbal Fluency Task

Category fluency and letter fluency: The participant's score in each task is the number of unique correct words in the given time (1 minute).

### Sustained Attention to Response Task (SART)

A participant makes a keypress to every number that appears as quickly as possible, apart from number 3.

Accuracy is simply 'key\_resp\_trial\_1.keys' compared to 'correctAns'. That will tell us when the participant pressed space when they shouldn't have (i.e., failed to withhold response to a '3').

Reaction time (RT) is simply the average of 'key\_resp\_trial\_1.rt'. This will give an average reaction time in seconds.

There are 5 blocks of 45 trials each = 225 trials (practice blocks are disregarded).

### Reading Span Task (RST)

RST combines a processing component (Lexical Decision: judging the correctness of sentences) and a storage component (memorising a series of words for a subsequent recall).

Lexical Decision accuracy is 'sentence\_resp.keys' compared to 'isValid' (column L). RTs are in 'sentence\_resp.rt' (this is just used to check that participants are paying attention and doing the task).

Recall must be calculated manually by comparing 'resp\_box.text' against 'Words' (column M) for each run. There are 12 runs with either 2,3,4 or 5 words to be remembered (3 for each length).

The main output measure is the recall proportion, so if someone recalls 2 words from a 4-word run, then that run is scored 0.5. Then proportions are simply averaged over all runs. Practice blocks are disregarded.

### N-Back Task

The Participant matches a letter to the 'target' (0-back) or indicates whether it matches the one presented 2 before (2-back).

Accuracy: Whether the response is correct (1) or not (0) is coded in column G (key\_resp\_2.corr) for 0-back and column AR (key\_resp\_3.corr) for 2-back.



RT: key\_resp\_2.rt for 0-back and key\_resp\_3.rt for 2-back.

Averages are calculated for 0-back and 2-back. Practice blocks are disregarded in this task.

### **Digit Span Task (DST)**

The participant does ‘forward’ then ‘backward’ span and then has to recall in ascending numerical order.

The number of digits increases 1 at a time (2 trials/span) until the participant fails on both trials.

Maximum span can be seen from Cols B for forward, AN for backwards, and AW for ascending, which indicate whether the participant entered the correct response. Maximum span is where they got 1 or both trials correct before failing both on the next increase.

### **Plus-Minus Task**

The participant completes a list of additions, followed by subtractions, and then a list where they must switch between adding and subtracting.

A switching cost is calculated where ‘switch cost’ = (time to complete switch condition) – (average of the addition/subtraction conditions).

## APPENDIX II: Scoring of the Interpersonal Skills Questionnaires

1. **State-Trait Anxiety Inventory (STAI)** (only trait version was used as part of the investigation)

- Bieling et al., (1998). The State-Trait Anxiety Inventory, Trait Version: Structure and Content Re-Examined. *Behaviour Research and Therapy*.
- There are 20 questions in total (trait version). STAI uses a 4-point scale where 1 = not at all and 4 = very much so. Reverse scoring is used for anxiety-absent items.

2. **Brief Resilience Scale (BRS)**

- Smith et al., (2008). The Brief Resilience Scale: Assessing the Ability to Bounce Back. *Int J Behav Med*.
- There are six questions (3 reverse scored) on a scale of 5 items.

3. **Barratt Impulsiveness Scale (BIS)**

- Meule et al., (2020). Confirmatory Factor Analysis of the Barratt Impulsiveness Scale–Short Form (BIS–15) in Patients with Mental Disorders. *Psychiatry Research*.
- 15 questions represent three subdivisions being measured: **A – Attention impulsivity**, **M – Motor impulsivity**, and **NP – Non-planning impulsivity**. The scores are from 1 = rarely to 4 = almost always. Non-impulsive statements are reverse scored.

4. **Cognitive Flexibility Scale (CFS)**

- Martin, M. M., & Rubin, R.B. (1995). A New Measure of Cognitive Flexibility. *Psychological Reports*.
- Participants are asked to answer 12 (6-point) Likert-type questions (from 1 = strongly disagree to 6 = strongly agree). The values of the 12 items are summed up.

5. **Personal Innovativeness in IT Scale (PIIT)**

- Lopez-Bonilla and Lopez-Bonilla (2012). Sensation-Seeking Profiles and Personal Innovativeness in Information Technology. *Social Science Computer Review*.
- Based on a 7-point Likert scale with the midpoint labelled neutral.



## 6. Ten Item Personality Indicators (TIPI)

- Gosling et al., (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality*.
- The TIPI consists of two items for each of the domains represented in the Five Factor Model (FFM) of personality (**Extraversion**, **Agreeableness**, **Conscientiousness**, **Emotional Stability**, and **Openness to New Experiences**).
- One item contains two desirable descriptors and the other, two undesirable descriptors (e.g., for Extraversion: extraverted, enthusiastic, and reserved and quiet).
- Each of the ten items are rated on a 7-point Likert scale from 1 (disagree strongly) to 7 (agree strongly).

## 7. The Work Extrinsic and Intrinsic Motivation Scale (WEIMS)

- Trembley et al., (2009). Work Extrinsic and Intrinsic Motivation Scale: Its Value for Organizational Psychology Research. *Canadian Journal of Behavioural Science*.
- WEIMS is an 18-item measure of work motivation theoretically grounded in self-determination theory.
- The items are rated on a 7-item scale from 1 = does not correspond at all to 7 = corresponds exactly.
- There are six three-item subscales (from the highest motivation to the lowest): **Intrinsic motivation** (4, 8, and 15), **Integrated regulation** (5, 10, and 18), **Identified regulation** (1, 7, and 14), **Introjected regulation** (6, 11, and 13), **External regulation** (2, 9, and 16), and **Amotivation** (3, 12, and 17).
- Work Self-Determination Index can be calculated (relative level of self-determination).

## 8. Five Facet Mindfulness Questionnaire (FFMQ), short version

- Bohlmeijer et al., (2011). Psychometric Properties of the Five Facet Mindfulness Questionnaire in Depressed Adults and Development of a Short Form. *Assessment*.
- The FFMQ (short version) has a total of 24 questions. The questions are scored on a 1-5 scale where 1 = never and 5 = very often. For the questions 4, 5, 7, 8, 11, 12, 14, 17, 19, 22, 23, and 24 the participant's score is subtracted from 6. The scores are then summed up based on the five different areas that are being measured: **Non react** (questions 3, 9, 13, 18, and 21); **Observe** (questions 6, 10, 15, and 20); **Act aware** (questions 8, 12, 17, 22, and 23); **Describe** (questions 1, 2, 5, 11, and 16); **Non-judge** (questions 4, 7, 14, 19, and 24).