SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Survey Practice Guide 6: Cognitive function measurement in online self-completion surveys

Cristian Domarchi[1], Olga Maslovskaya[1], Lisa Calderwood[2], Matt Brown[2]
[1] University of Southampton, [2] University College London

February 2026

www.surveyfutures.net

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefiting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net

Research Strand 5 of Survey Futures ("Complex measurements"), led by Professor Lisa Calderwood (UCL), focuses on the challenges associated with administering complex measures in online surveys without detriment to data quality and/or comparability. Research Strand 5 focuses on four types of complex measures:

(1) Industry and occupation coding.
(2) Consent to data linkage.
(3) Retrospective data collection.
(4) Cognitive function measurement.

The suggested form of citation for this guide is:

Domarchi C, Maslovskaya O, Calderwood L & Brown M (2025) 'Cognitive function measurement in online self-completion surveys', *Survey Futures Survey Practice Guide* no. 6. Colchester, UK: University of Essex. Available at https://surveyfutures.net/practice-guides/.

Contents

Introduction

Cognitive function is a key determinant of health, economic, educational, and social outcomes across the life course. Incorporating measures of cognition into large-scale surveys enables researchers to track population trends, detect early signs of decline, and design and evaluate interventions. However, administering cognitive assessments in surveys is challenging. Many instruments, especially those which were originally developed for use in medical or psychological clinics, are designed for in-person administration. As surveys increasingly transition to online data collection, this creates significant difficulties, and there is strong evidence that cognitive assessments are particularly vulnerable to mode effects.

This guide summarises different methods used to measure cognitive function in online self-administered surveys, and provides recommendations for survey practice, based on our evidence review (Domarchi *et al.*, 2025). We use evidence from both survey practice and academic literature about measurement approaches, assessing their advantages and limitations. We first describe the primary approaches used and then make some general recommendations for survey practice. The detailed findings and recommendations are then set out in a series of summary tables. First, Table 1 briefly summarises the pros and cons of each method. The following tables (Tables 2 to 6) describe each approach along with their advantages and limitations, and specific practical considerations for implementation. They also provide some general recommendations for method selection based on survey type and data collection requirements. Finally, Table 7 provides more detail about some of the more commonly used test batteries which have been used to measure cognition in large-scale online surveys, including both adapted batteries (type 1), and batteries specifically designed for online self-administration (type 2). The table provides a brief description of each test, along with the cognitive domain covered, and the surveys that have used each.

Approaches to measuring cognition in online surveys

Our evidence review identifies **five primary approaches** to measuring cognition in online self-administered surveys. These approaches are distinguished by their data collection methods:

1. **In-person assessments adapted for online self-administration.** Several long-running longitudinal surveys, such as the Health and Retirement Study (HRS) and Understanding America (UAS) in the US, and Understanding Society in the UK, have historically collected cognitive data through interviewer-administered modes (e.g., face-to-face or telephone interviews). As surveys increasingly move online, whether in single-mode designs or as part of mixed-mode strategies, numerous studies have adapted tests originally designed and validated for interviewer-led administration for use in web-based settings.
2. **Assessments specifically designed for online self-administration.** More recently, longitudinal surveys have been using cognitive batteries (i.e. standardised sets of cognitive tests measuring different aspects of cognition) specifically designed for

online and remote self-administration. Widely validated test batteries used in large-scale social surveys include Cognitron (used in the National Survey of Health and Development, NHSD, in the UK), CogState (used in the Raine Study in Australia), TestMyBrain (used in the Twins Early Development Study TEDS in the UK), and the UK Biobank cognitive test battery.

3. **Measures derived from survey response behaviour.** Recent research suggests that examining survey response behaviour may represent a valuable resource that can be used to develop behaviour-based markers of cognitive decline that are cost-effective, unobtrusive, and scalable. Specifically, indicators of low response quality or low response speed, have been linked to cognitive decline, especially in the long term, using survey data and paradata from longitudinal studies including HRS, the English Longitudinal Study of Ageing (ELSA), and the Survey of Health, Ageing and Retirement in Europe (SHARE).
4. **Measures derived from apps and games.** Gamification (or the integration of game design elements into non-games contexts), has been more recently used to measure cognitive abilities in several social surveys such as the SeaHeroQuest app in Understanding Society, and Pathfinder in TEDS.
5. **Measures obtained through mobile technologies.** The widespread use of smartphones (including among the ageing population) with increasing storage capacity and connectivity, has enabled the possibility of collecting large amounts of data with minor effort required from participants. Smartphone-based cognitive assessments include mobile versions of existing tests, and new tests for mobile devices, which are usually implemented as brief, frequent, and repeated assessments. In addition, other data types collected via smartphone or sensor apps including speech analysis, physical movement analysis, GPS data, and activity daily-life performance have been used in many studies as indicators of cognitive ability.

General recommendations for survey practice

Our key overall recommendations are:

- **Prioritise cognitive measures specifically designed for online self-completion rather than attempting to adapt assessments designed to be conducted in-person.** In mixed-mode surveys which involve both web and in-person interviews, those participating in-person should ideally complete the online assessment via self-completion. This approach has been shown to reduce (or even eliminate) mode effects.
- For studies transitioning from in-person interviewing to web-based (or web-first) data collection, **conduct calibration studies where sub-samples of participants complete both the legacy in-person assessments and the new online measures**. Although logistically demanding and costly, such studies may be beneficial for safeguarding data comparability across survey waves and providing evidence on mode differences for adjustment.

Additionally, some practical recommendations for the implementation of cognitive assessments in online surveys include:

- Allow participants to pause between tests or complete the assessments over multiple days, with easy re-access to the testing platform, to reduce potential stress.
- Facilitate access by providing detailed video or written instructions and hosting both the consent process and assessments within the same system. A detailed frequently asked questions (FAQ) page, task-specific instructions, practice trials, and interactive tutorials are also recommended.
- Notify participants that a stable internet connection is required. Provision of a phone or email helpline to address potential technical issues is recommended.
- Design the test interface with clear language, intuitive icons, and a simplified consent form to accommodate respondents with lower technical literacy.
- Use larger font sizes, limit flashing lights, and offer an audio option for written instructions. These steps can facilitate participation by elderly users.

Summary of methods

Table 1. Summary of pros and cons of each method

Method	Pros	Cons
In-person assessments adapted for online self-administration	• Maintains longitudinal consistency with in-person tests from previous survey waves. • Self-administration reduces the potential for interviewer-led biases. • Benefits from well-established psychological validity and reliability.	• Strong evidence of mode effects, particularly for respondents with lower cognitive ability. • Unsupervised administration can compromise data quality. • Device type, screen size, and internet speed can influence performance, introducing new biases.
Assessments specifically designed for online self-administration	• Removes potential interviewer-led biases • Built on a foundation of well-established validity and reliability • Helps minimise mode effects.	• Unsupervised administration can compromise data quality • Device type, screen size, and internet speed can influence performance, introducing new biases.
Measures derived from survey response behaviour	• Non-intrusive and cost-effective • Provides direct observation of a complex, real-world cognitive task • Paradata are often readily available in existing surveys • Can assess multiple cognitive domains simultaneously	• Response behaviour can be influenced by factors confounded with cognitive ability, such as: ○ Participant motivation and engagement ○ Social desirability bias and survey fatigue ○ Device type, connectivity, and questionnaire design • Reliability and validity of these indicators as proxies for cognitive ability are not yet well-established
Measures derived from apps and games	• Increases participant engagement and reduces attrition • Can provide a more authentic measure than static tests by simulating real-world dynamic environments • Generates rich, process-oriented data beyond simple performance scores.	• Game performance can be influenced by factors confounded with cognitive ability, such as: ○ Participant motivation ○ Familiarity with game mechanics ○ Device type and connectivity ○ General aptitude for digital games ○ Design and implementation can be costly • Reliability and validity of game performance as a proxy for cognitive ability are not yet well-established
Measures obtained through mobile technologies	• Non-intrusive and requires minimal user effort • Collects large volumes of high-frequency data • Increases participant engagement and improves accessibility • Reduces or eliminates interviewer effects	• Performance can be influenced by factors confounded with cognitive ability, including: ○ Contextual and environmental conditions ○ User's physical state ○ Device type and connectivity • Risk of attrition and missing data due to battery consumption, privacy concerns, and the need to carry the device • Reliability and validity of these indicators as proxies for cognitive ability are not yet well-established

Table 2. In-person assessments adapted for online self-administration

Summary / Examples	Pros	Cons	Wider considerations	Recommendations
Established longitudinal surveys, including the HRS in the US and Understanding Society in the UK, have adapted in-person cognitive tests for online self-administration in mixed-mode survey contexts. Test have been extracted from widely validated batteries including: • The Mini-Mental State Examination (MMSE). • The Weschler Adult Intelligence Scale-Revised. • The Woodcock-Johnson Tests of Cognitive Ability. Further details about these tests can be found on Table 2.	• In the context of longitudinal studies switching from in-person to web (or mixed mode), the time-series can be maintained by keeping longitudinal consistency with in-person tests used in previous waves • Can remove interviewer-led biases, allowing for more uniform administration across respondents. • The psychological validity and reliability of these batteries are well-established in the academic literature.	• Strong evidence shows that adapting face-to-face tests for online use introduces mode effects, especially among respondents with lower cognitive functioning, potentially distorting substantive findings. • Unsupervised administration can compromise data quality, as participants may misunderstand instructions, be easily distracted, or not apply full effort. • Device type, screen size, and internet speed can influence test performance, introducing bias unrelated to cognitive ability. • Some tasks may be too complex, making them unsuitable for unsupervised online formats	• The scientific validity and reliability of any chosen battery must be confirmed, and tests should be sensitive enough to detect subtle differences within the target cognitive domain. • Ease of administration is crucial for minimising breakoffs and incomplete responses. • User testing is essential for defining layout and design parameters, ensuring the tool is accessible for the entire target population. • The costs of licensing established batteries or developing new methods to collect cognitive data can significant.	• Adapting batteries originally designed for face-to-face administration in mixed-mode surveys including self-administration is not recommended, as strong mode effects are expected. • In the context of mixed-mode surveys including self-administration, in-person participants should complete cognitive assessments via self-completion, to ensure some degree of comparability with self-administered online assessments. However, in this case, longitudinal mode effects may still be an issue if in prior waves tests were interviewer administered • When moving from in-person administration to mixed mode surveys including self-administration, calibration studies may be required for quantifying mode effects and testing adjustment strategies. In these studies, participants complete both the legacy measures and the newly introduced ones. • To ensure response quality, simpler tasks with short responses should be prioritised over complex ones that are highly dependent on specific devices. Instructions should be simple and accessible, and test batteries should feature large fonts, audio support, clear language, and intuitive icons. • Participants should be encouraged to complete the assessments in a quiet environment to minimise potential distractions. • Participants should be reminded that honest answers are essential and that a perfect score is not expected.

Table 3. Assessments specifically designed for online self-administration

Summary / Examples	Pros	Cons	Wider considerations	Recommendations
Several longitudinal surveys measure cognitive ability using batteries specifically designed for online self-administration. Some examples include: • Tests from the *Cognitron* online battery were used in the 1946 National Birth Cohort (UK). • Tests from the *Cogstate* battery were used in the Raine Study and the 45 and Up Study (both in Australia). • Tests from the *TestMyBrain* battery were used in the Millennium Cohort Study, Next Steps, and the Twins Early Development Study (TEDS) in the UK. • The *UK Biobank* survey has developed its own battery of cognitive test batteries. Further details about these tests can be found on Table 2.	• Can remove interviewer-led biases, allowing for more uniform administration across respondents. • The psychological validity and reliability of these batteries are well-established in the academic literature. • Using test batteries designed for online self-administration in mixed-mode surveys helps to minimise mode effects.	• In the context of longitudinal studies switching from in-person to web (or mixed mode) this may require a break in time-series as new tests may need to be introduced • Unsupervised administration can compromise data quality, as participants may misunderstand instructions, be easily distracted, or not apply full effort. • Device type, screen size, and internet speed can influence test performance, introducing bias unrelated to cognitive ability.	• Age group is a key consideration for the selection of a test battery, as cognitive ability, attention span, and technological familiarity vary significantly across different life stages. • The scientific validity and reliability of any chosen battery must be confirmed, and tests should be sensitive enough to detect subtle differences within the target cognitive domain. • Ease of administration is crucial for minimising breakoffs and incomplete responses. • User testing is essential for defining layout and design parameters, ensuring the tool is accessible for the entire target population. • The costs of licensing established batteries or developing new methods to collect cognitive data can be significant. • Ensuring data confidentiality is crucial to protect participant privacy and comply with legal and ethical standards.	• Administration of batteries specifically designed for online self-administration should be the preferred option, as it reduces mode effects and helps ensure more robust cognitive scores. • In the context of mixed-mode surveys including self-administration, in-person participants should complete cognitive assessments via self-completion, to ensure some degree of comparability with self-administered online assessments. • When moving from in-person administration to mixed mode surveys including self-administration, calibration studies may be required for quantifying mode effects and testing adjustment strategies. In these studies, participants complete both the legacy measures and the newly introduced ones. • To ensure response quality, simpler tasks with short responses should be prioritised over complex ones that are highly dependent on specific devices. • Participants should be encouraged to complete the assessments in a quiet environment to minimise potential distractions. • Participants should be reminded that honest answers are essential and that a perfect score is not expected.

Table 4. Measures derived from survey response behaviour

Summary / Examples	**Pros**	**Cons**	**Wider considerations**	**Recommendations**
Recent research explores how cognitive ability can be inferred by assessing survey response behaviour. Two types of indicators from longitudinal surveys have been studied: • *Response quality indicators:* o item non-response (skipped questions) and uncertainty (proportion of "don't know" answers") o response styles (acquiescence, dis-acquiescence, extreme and midpoint responses, straight-lining, pattern drift) o response consistency (Guttman errors, multivariate outliers, improbable responses) • *Survey para-data indices:* o survey completion (time lag from invitation to uptake, time lag from uptake to completion, missed survey) o survey response time o errors and corrective behaviours (e.g. corrected or changed answers, error messages received, use of "back" button, use of "next" button without a complete response), o mouse and touch efficiency (number of clicks per page, total pixel count) o keystrokes (time between keystrokes).	• Survey response behaviour can provide valuable indicators of cognitive functioning in a non-intrusive and cost-effective manner, compared to conventional cognitive assessments. • It enables the direct observation of a complex, real-world cognitive task (the act of survey completion) without dedicated testing, making it particularly well-suited for large-scale population studies. • This type of data is often readily available in many longitudinal datasets, facilitating comparative research across studies. • Since responding to a survey demands cognitive skills such as attention, working memory, executive function, and both short- and long-term memory, these behavioural indicators can offer detailed information about multiple cognitive domains simultaneously.	• Survey response behaviour is influenced by numerous factors beyond cognitive ability, such as participant motivation and engagement, social desirability bias, survey fatigue, device type and connectivity and questionnaire design. These confounding variables can complicate the interpretation of behaviour as a direct indicator of cognitive function. • The psychological validity and reliability of using survey response patterns as proxies for cognitive ability have not yet been firmly established, as this area of research is still emerging.	• While initial evidence suggests an association between survey response behaviour and cognitive ability (particularly over the longer term), the strength of this relationship is weak.	• Indicators derived from survey paradata show no indication of being able to fully replace conventional cognitive assessments in terms of reliability or diagnostic accuracy. • Instead, survey response behaviour is best positioned as a supplementary tool to increase understanding of cognitive ability decline over time, screen for potential issues, or identify candidates for more in-depth evaluation.

Table 5. Measures derived from apps and games

Summary / Examples	Pros	Cons	Wider considerations	Recommendations
Cognitive ability has been assessed with games using two primary approaches: specially designed games created to boost engagement, and conventional cognitive tests embedded within a gamified environment. Examples from UK surveys include: • *Sea Hero Quest* (Understanding Society Innovation Panel): A smartphone and tablet-based video game that assesses wayfinding and path integration. Players navigate a boat from a starting point to various destinations on a map. • *Pathfinder* (TEDS): A gamified assessment that embeds five cognitive tests (visual puzzles, matrix reasoning, verbal analogies, vocabulary, and missing letters) within an engaging storyline.	• Games can enhance participant engagement and reduce attrition by making cognitive tasks more enjoyable and motivating. • When games simulate real-world challenges in dynamic environments, they can provide a more authentic measure of cognitive processes than static test items. • Games have the potential to generate rich, process-oriented data beyond simple performance scores, offering more detailed information about latent cognitive constructs.	• The psychological validity and reliability of using game performance indicators as proxies for cognitive ability have not yet been firmly established, as this area of research is still emerging. • Game performance is influenced by numerous factors beyond cognitive ability, such as participant motivation, familiarity with game mechanics, device type or connectivity or general affinity for digital games. These confounding variables can complicate the interpretation of game performance as a direct indicator of cognitive function. • The design and implementation of purpose-built cognitive games are typically more cost-intensive than deploying a conventional test battery.	• Early evidence from *Sea Hero Quest* suggests that game performance measures (e.g., reaction time, error rate) are correlated with cognitive abilities, particularly in domains such as processing speed, executive control, and task-switching. • Measures derived from cognitive tests embedded in a gamified storyline (*Pathfinder*) demonstrated reliable verbal and non-verbal scores, which correlated substantially with standard cognitive measures collected at earlier ages in the study.	• Measures derived from games may serve as a valuable supplementary tool to traditional cognitive test batteries, particularly for enhancing our understanding of cognitive function or for identifying candidates for further evaluation. • Games can be particularly effective for engaging younger respondents, a demographic often characterized by lower survey participation rates.

Table 6. Measures obtained through mobile technologies.

Summary / Examples	Pros	Cons	Wider considerations	Recommendations
The widespread use of smartphones (including among the ageing population) with increasing storage capacity and connectivity has enabled the possibility of collecting a wide variety of data through apps which can be used as indicators of cognitive ability data including: ○ apps collecting GPS data in smartphones and wearable sensors ○ activity daily-life performance apps ○ speech analysis apps ○ physical movement apps Mobile versions of existing tests, and new tests for mobile devices, usually implemented as brief, frequent, and repeated assessments. There have also been early studies linking cognitive ability with interaction between participants and smart home devices. These mobile applications have not yet been used in large-scale probability-based surveys.	• Smartphones can collect large volumes of high-frequency, longitudinal data in a non-intrusive manner, requiring minimal effort from participants. • Applications can increase participant engagement, reduce interviewer effects, and improve accessibility. This can lead to larger, more inclusive sample sizes and a corresponding reduction in selection bias.	• Performance on app-based assessments is influenced by numerous factors beyond cognitive ability, including contextual and environmental conditions, physical state, and technical issues related to the device or internet connection. This complicates data interpretation and introduces a risk of confounding. • Passive data collection via apps involves gathering highly sensitive information, which requires obtaining explicit informed consent and implementing secure protocols for data storage and transmission. • Practical challenges such as battery consumption, privacy concerns, and the need to continuously carry the device can lead to higher attrition rates and increased missing data. • The psychological validity and reliability of cognitive measures derived from apps have not yet been sufficiently established, as this remains an emerging area of research. • The development of specialized smartphone applications can be cost-intensive.	• Evidence suggests that apps and passive data collection can generate meaningful measures of cognitive ability. Key findings include: ○ the total distance covered, as measured by GPS apps, has been shown to be significantly lower for older adults with cognitive impairments compared to their healthy counterparts. ○ performance in apps that simulate daily activities (e.g., a virtual supermarket) has been reliably linked to cognitive ability. ○ various forms of dementia and mild cognitive impairment can manifest as detectable irregularities in speech and language patterns. ○ significant differences in finger dexterity have been observed between individuals with dementia and healthy control populations. • However, significant barriers to widespread adoption remain, particularly in obtaining participant consent and ensuring sustained engagement with the application.	• Measures derived from innovative devices and applications should be considered a supplementary tool to traditional cognitive test batteries. They are particularly valuable for enhancing the understanding of cognitive function and for identifying individuals who may require further clinical evaluation. • Applications designed to detect and analyse subtle changes in speech and physical movement show significant promise as tools for the early detection of cognitive diseases such as dementia and Alzheimer's.

Description of tests

The following table lists the cognitive ability test batteries used in the large-scale probability-based surveys reviewed. The table includes the following fields:

- The "**type**" column classifies each test battery as (1) adapted for online administration or (2) designed for online self-administration
- The "**survey**" column indicates the survey in which each test was used. Survey abbreviations are as follows: HRS = Health and Retirement Survey, US-IP = Understanding Society Innovation Panel, UAS = Understanding America Survey, PSID-WEB: Panel Study of Income Dynamics - Well Being and Daily Life Supplement, NHSD = National Survey of Health and Development (1946 Birth Cohort Study), TEDS = Twins Early Development Study.
- The "**domain**" column indicates the cognitive domain measured by each test.
- The "**test name**" column contains the name of each test. In addition, it provides a link that directs to a webpage, technical report, or academic paper in which the test is described in more detail.
- The "**summary**" column provides a brief description of each test.

Table 7. Description of cognitive tests implemented in surveys.

Type	Battery	Surveys	Domain	Test name	Summary
1	HRS	HRS, US-IP, UAS	Working memory	Immediate word recall	In the "immediate recall" question, the interviewer reads a list of 20 nouns to the respondent and asks the respondent to recall as many words as possible from the list in any order.
1	HRS	HRS, US-IP, UAS	Working memory	Delayed word recall	In the "delayed" version of the test, respondents are asked to recall the nouns previously presented as part of the immediate recall task (5-7 minutes ago).
1	HRS	HRS, US-IP, UAS	Executive function	Backwards count	Respondents are asked to count backwards for 10 continuous numbers beginning with the number 20 (and from 86 in some versions).
1	HRS	HRS, US-IP, UAS	Executive function	Serial 7's test	The interviewer asks the respondent to subtract 7 from 100 and continue subtracting 7 from each subsequent number for a total of five trials. It was up to the respondent to remember the value from the prior subtraction, such that the interviewer did not repeat the difference said by the respondent after each trial.
1	HRS	HRS, UAS	Orientation	Date naming	Respondents are asked to report "today's date".
1	HRS	HRS, UAS	Language/Naming	Object naming	Questions include: "What do you usually use to cut paper?" and "What do you call the kind of prickly plant that grows in the desert?".

Type	Battery	Surveys	Domain	Test name	Summary
1	HRS	HRS, UAS	Language/Naming	President/Vice-president naming	Respondents are asked to name the current President and Vice President of the United States.
1	HRS	HRS	Vocabulary	Vocabulary test (adapted from WAIS-R)	Respondents are asked to define 5 words from a closed list.
1	HRS	HRS, US-IP	Numeracy	Questions to measure numeric ability	Three simple arithmetic problems are given to respondents to solve.
1	HRS	HRS	Fluid intelligence	Animal naming	Participants are asked to name as many animals as they can in 60 seconds.
1	HRS	HRS, US-IP, UAS	Fluid intelligence	Verbal analogies	Participants are given six verbal analogies to complete (e.g. "mother is to daughter as father is to […]".
1	HRS	HRS, PSID-WEB	Fluid intelligence	Number series test adapted from the Woodcock-Johnson (WJ-R) tests of cognitive ability	Participants are given a series of numbers with a blank space to be completed.
1	HRS	PSID-WEB	Language competence	Verbal reasoning test	Participants are given a series of sentence completion questions. They are asked to select the word that makes the most sensible complete sentence.
2	Cognitron	NHSD	Working memory	Objective immediate and delayed recognition	Participants are shown a sequence of target objects. They are asked to identify these targets in different arrays of objects. Like the HRS delayed word recall, the task is repeated at the end of the battery to measure delayed memory recognition.
2	Cognitron	NHSD	Processing speed	Motor control	Participants are shown a red target appearing at different locations of the screen and asked to tap on it as quickly as possible.
2	Cognitron	NHSD	Processing speed	Choice reaction time	Participants are shown an arrow pointing either left or right and must respond accordingly to it tapping on the left or right-hand of the screen.
2	Cognitron	NHSD	Visuospatial abilities	Blocks	Participants are asked to remove blocks of different colours and shapes from one array to match the target array.
2	Cognitron	NHSD	Executive function	Digit span	Participants are asked to memorise a list of digits and then repeat it. The list of digit increases in length every correct trial. The task is interrupted after three consecutive incorrect trials.
2	Cognitron	NHSD	Visuospatial abilities, Attention/Concentration	Spatial span	Participants are asked to memorise a sequence of grey squares appearing at different locations of a 4x4 grid. The number of squares increases in length every correct trial. The task is interrupted after three consecutive incorrect trials.
2	Cognitron	NHSD	Executive function	Stroop	Participants indicate the colour of a title by tapping "blue" or "red", which are coloured either blue or red. A box indicates the modality they will have to provide the answer in (colour or text of the word).

Type	Battery	Surveys	Domain	Test name	Summary
2	Cognitron	NHSD	Visuospatial abilities	2-D manipulations	Participants are shown a target array of coloured squares and asked to identify this among four. The target is rotated through either 90, 180, or 270 degrees.
2	Cognitron	NHSD	Language competence	Word definitions	Participants are shown a word and 4 possible definitions and asked to tap on the correct definition within a designated amount of time.
2	Cognitron	NHSD	Language competence	Verbal reasoning	Participants are shown different combinations of geometric shapes and asked to indicate whether the statement describing the shapes is true or false.
2	Cognitron	NHSD	Processing speed	Spotter	Participants see numbers displayed inside a pixelated square. They are asked to click on the square immediately upon spotting the number "0". The stimuli appear on the screen for only 100 ms, in rapid succession, and are degraded with a mask.
2	Cognitron	NHSD	Processing speed	Forager	Participants see a continuous stream of shapes. They are asked to click on the shapes until they find the correct rule (e.g. tap on circles). They will do so based on the feedback they receive (correct/incorrect). After they follow the rule correctly for 6 consecutive trials, they receive negative feedback and a new rule is generated (e.g., tap on squares).
2	UK Biobank	UK Biobank	Visual declarative memory	Pairs matching test	Participants are asked to memorise the position of as many matching pairs of cards as possible. The cards are then turned face down on the screen and participants are asked to touch as many pairs as possible in the fewest tries.
2	UK Biobank	UK Biobank	Processing speed	Reaction time test	Participants are shown two cards at a time; if both cards are the same, they press a button-box on the table in front of them as quickly as possible.
2	UK Biobank	UK Biobank	Prospective memory	Prospective memory test	Early in the touchscreen cognitive section, participants are shown the message "At the end of the games we will show you four coloured shapes and ask you to touch the Blue Square. However, to test your memory, we want you to actually touch the Orange Circle instead."
2	UK Biobank	UK Biobank	Verbal and numerical reasoning	Fluid intelligence test	Participants have 2 minutes to complete as many questions as possible from the test.
2	UK Biobank	UK Biobank	Working memory	Numeric memory test	Participants were shown a 2-digit number to remember. The number then disappeared and after a short while they were asked to enter the number onto the screen. The number became one digit longer each time they remembered correctly (up to a maximum of 12 digits).
2	UK Biobank	UK Biobank	Executive function	Trail making test parts A and B	Participants were presented with sets of digits/letters in circles scattered around the screen and asked to click on them sequentially according to a specific algorithm.
2	UK Biobank	UK Biobank	Processing speed	Symbol digit substitution test	Participants were presented with one grid linking symbols to single-digit integers and a second grid containing only the symbols. They were then asked to indicate the numbers attached to each of the symbols in the second grid using the first one as a key.
2	UK Biobank	UK Biobank	Vocabulary	Picture vocabulary	Participants were presented with a series of sets of four pictures accompanied by a word and asked to indicate which image in the set was most closely related to the word displayed.
2	UK Biobank	UK Biobank	Verbal declarative memory	Paired associate learning test	Participants were shown 12 pairs of words (for 30 seconds in total) then, after an interval (in which they did a different test), presented with the first word of 10 of these pairs and asked to select the matching second word from a choice of 4 alternatives.

Type	Battery	Surveys	Domain	Test name	Summary
2	UK Biobank	UK Biobank	Executive function	Tower rearranging test	Participants were presented with an illustration of three pegs (towers) on which three differently coloured hoops had been placed. The were then asked to indicate how many moves it would take to re-arrange the hoops into another specific position.
2	UK Biobank	UK Biobank	Non-verbal reasoning	Matrix pattern completion	Participants were presented with a series of matrix pattern blocks with an element missing and asked to select the element that best completed the pattern from a range of displayed choices.
2	Cogstate	Raine Study	Visuospatial abilities, Working memory	Continuous Paired Associate Learning Test (CPAL)	Participants must learn and remember pictures hidden beneath different locations on the screen.
2	Cogstate	Raine Study	Processing speed	Detection Test (DET)	This test measures processing speed using a simple reaction time paradigm ("has the card turned over"?)
2	Cogstate	Raine Study	Executive function, Administration	Groton maze learning test (GMLT)	A 10 x 10 grid of tiles is presented on the screen, with a 28-step pathway hidden among these tiles. Participants must move one step at a time from the start toward the end by touching a tile next to their current location. Once completed, they return to the start to repeat the test, trying to remember the pathway they just completed.
2	Cogstate	Raine Study	Attention/Concentration	Identification Test (IDT)	A key is provided at the top of the screen pairing nine medicines with a date. In the middle of the screen, an empty pill box labelled with a date is presented and the subject is asked to select the medicine that corresponds to that date, as per the key.
2	Cogstate	Raine Study	Working memory	One Card Learning Test (OCL)	A playing card is presented face up in the centre of the screen and participants must decide whether they have seen it before in the test.
2	Cogstate	Raine Study	Working memory	One Back Test	A playing card is presented face up in the centre of the screen. Participants must decide whether the card is the same as the previous card.
2	Cogstate	Raine Study	Executive function	Set-shifting test	Participants are shown a playing card on a screen, accompanied by the word "Number" or "Colour," which determines the target rule. Based on this cue, participants guess if the card matches the target attribute (e.g., colour: black/red or number correctness) by pressing "Yes" or "No". Feedback is provided after each guess, and progression requires correct responses. Periodically, the target rule changes without warning (either within the same dimension or across dimensions), requiring participants to deduce the new rule.
2	TEDS	TEDS	Non-verbal reasoning	Ravens Progressive Matrices	A series of incomplete patterns ("matrices"). Participants are asked to identify the missing part on each pattern. They do so by clicking on one of 8 possible missing parts. The full battery (60 items) was presented at age 10. At age 12, participants responded to 24 items, while at ages 14 and 16, they responded to 30 items.
2	TEDS	TEDS	Non-verbal reasoning	WISC: Picture completion	A series of pictures of recognisable objects or scenes, each with an essential detail omitted from the picture. Participants are asked to click on the relevant part of the picture on screen, to identify the part that is missing. There is a time limit of 20 seconds for each picture. There were 30 items.
2	TEDS	TEDS	Vocabulary	WISC: Vocabulary	The test consisted of a series of vocabulary questions. For each question, there were either three or four possible responses; participants select a response by clicking on it on the screen. There were 30 items.

Type	Battery	Surveys	Domain	Test name	Summary
2	TEDS	TEDS	General knowledge	WISC: General knowledge	The test consists of a series of general knowledge questions. For each question, there are four possible responses; participants select a response by clicking on it on the screen. There are 30 items,
2	TEDS	TEDS	General knowledge	Author recognition	A list of 42 author names is presented on screen. 21 are names of real authors, and 21 are dummy names. Participants are asked to select all the real author names.
2	TEDS	TEDS	Reading ability	Woodcock-Johnson III Reading Fluency	Up to 98 yes/no statements, within a 3-minute time limit (2.5-minute limit for Age 16). Participants need to indicate "yes" or "no" ("true" or "false") for each statement, as quickly as possible.
2	TEDS	TEDS	Reading ability	GOAL Formative Assessment in Literacy (Key Stage 3)	Multiple choice of reading comprehension ability.
2	TEDS	TEDS	Numeracy	Number games	The test included 3 sub-tests: understanding numbers (33 items), non-numerical processes (25 items), and computation and knowledge (37 items).
2	TEDS	TEDS	Language competence	Test of language competence (Expanded edition). Semantics: The figurative language subset	Participants were asked to match expressions (or figures of speech) having similar meanings. There were 11 items in the test (15 items for Age 16). A "situation" is first played from an audio recording, followed by an "expression" (figure of speech), and a list of four new expressions, which are the response options. There is a 60-second limit for each response.
2	TEDS	TEDS	Language competence	Test of language competence (Expanded edition). Pragmatics: The making inference subset	Participants were asked to make inferences about the possible causes of given situations. There were 11 items in the test. An initial statement (two sentences) is played on an audio recording, following by a question, and a list of four response options. There is a 70-second limit for each response.
2	TEDS	TEDS	Language competence	Listening grammar subtest of the TOAL-3	A series of 35 items in which participants had to select two sentences with similar (or identical) meanings, from three sentences played using audio recordings
2	TEDS	TEDS	Vocabulary	Mill Hill Vocabulary test	A series of 33 multiple-choice questions. In each question, a single word is presented at the top of the screen. Below it, 6 other words are presented as the response options, with participants asked to click on the option they think is closest in meaning to the word at the top of the screen.
2	TEDS	TEDS	Reading ability	Passages comprehension	Participants are presented with two passages of written text, each of which is followed by 13 multiple-choice comprehension questions based on the text. The relevant passage remains on screen to allow twins to re-read it if necessary

Type	Battery	Surveys	Domain	Test name	Summary
2	TEDS	TEDS	Visuospatial abilities	Hidden shapes test	Each of the 27 items of this test displays a geometric shape which was hidden within one of four more complex patterns, also displayed on screen. Participants must decide which of the four patterns concealed the given shape.
2	TEDS	TEDS	Visuospatial abilities	Jigsaws tests	Each of the 27 items of this test displays four shapes, one of which has been divided into several jigsaw pieces. Participants had to decide which of the four shapes matches the assembled jigsaw pieces.
2	TEDS	TEDS	Visuospatial abilities	Eyes test	A series of 28 photographs of the eye region of the face of different actors and actresses is presented. In each case, participants were asked to choose which of four words (e.g., "jealous", "scared", "relaxed", "hate") best describes what the person in the picture was thinking or feeling.
2	TEDS	TEDS	Science, General knowledge	Science test	A test of Scientific Enquiry skills, based on the UK National Curriculum. The test included 39 items.
2	TEDS	TEDS	Visuospatial abilities	Bricks test	Mental rotation and visualisation were measured separately and together, using both 2D and 3D stimuli, to form a battery of six subtests. Each of the 12 items presents a target stimulus image and four response images, with participants asked to select the response image showing the same object as shown in the target.
2	TEDS	TEDS	Visuospatial abilities	Kings Challenge test	A battery of 10 activities to test spatial abilities. The activities included: cross-section test, 2D drawing test, pattern assembly test, Elithorn mazes test, mechanical reasoning test, paper folding test, 3D drawing test, shape rotation test, perspective taking test, and Mazes test.
2	TEDS	TEDS	Visuospatial abilities	Navigation web study	A set of 30 related game-like activities to test navigational spatial abilities. The game included activities measuring spatial orientation, map reading, scanning, and perspective. A shortened version was used at Age 26.
2	Test My Brain	TEDS	Vocabulary	Vocabulary test	On each trial, participants selected which of five response option words are closest in meaning to a probe word. The standard length, hard version of the test contained 20 test trials.
2	Test My Brain	TEDS	Processing speed	Digit Symbol Matching test	Participants were presented with six symbols, each of which is paired with a single digit between 1-3 (i.e., two symbols are paired with each digit). These digit-symbol pairings remained visible throughout the duration of the test. Individual probe symbols are sequentially presented above the digit-symbol pairings, to which participants responded by selecting the corresponding digit as quickly as possible.
2	Test My Brain	TEDS	Verbal declarative memory	Verbal Paired Associates Memory test	Participants were visually presented with 25 pairs of words and informed they will later be tested on which words were paired together. After a delay of approximately 1.5-2.5 minutes, during which another brief test was typically completed, participants were sequentially presented with one word from each of the studied word pairs, and asked to identify which word was previously paired with it by selecting the correct word from a list of four response options
2	Test My Brain	TEDS, Next Steps, MCS	Executive function	Forward and backward digit span test	After being presented with a set of numbers, participants were asked to recall those numbers either in their original order (for the forward test), or in reverse order (for the backward test).

Reference

Domarchi, C., Maslovskaya, O., Calderwood, L. & Brown, M. (2025) Cognitive function measurement in online self-completion surveys: evidence review. *Survey Futures Report No. 9*. Colchester, UK: University of Essex. Available at: https://surveyfutures.net/reports /