SURVEY FUTURES
SURVEY DATA COLLECTION METHODS COLLABORATION

Report 9: Cognitive function measurement in online self-completion surveys: Evidence review

Cristian Domarchi[1], Olga Maslovskaya[1],
Lisa Calderwood[2], and Matt Brown[2]

[1]University of Southampton, [2]University College London

November 2025

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 (University College London), 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 report is:

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/.

Contents

Executive summary

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 evaluate interventions. However, administering cognitive assessments in surveys is challenging. Many instruments are adapted from integrated sets of standardised psychometric assessments (or *test batteries*) used in a clinical context. These are typically designed for in-person administration. As surveys increasingly transition to web-based data collection, this creates significant difficulties, and there is strong evidence that cognitive assessments are particularly vulnerable to mode effects.

This evidence review synthesises experiences from surveys from around the world that have implemented web-based cognitive measures across different populations, with a particular focus on large-scale probability-based online surveys from the UK. The review examines the cognitive assessments and test batteries used for online self-administration, highlighting their strengths, limitations, and practical considerations. It also explores innovative approaches from outside traditional surveys, including novel methods, tools, and devices for online cognitive measurement.

The review demonstrates that cognition has been assessed in online surveys using conventional test batteries (generally developed for in-person surveys) adapted for online administration, batteries specifically designed for online administration, or app-based assessments, with valid and reliable tests available for all these options. In addition, indicators collected in surveys via paradata, including for example item non-response and response times, can offer further insights into cognitive function.

However, substantial challenges remain. Selecting an appropriate test battery depends on the research objectives and the cognitive domains to be assessed. Key considerations include the capabilities of each survey mode, the target age group, the scientific validity and reliability of the tests, practical administration issues, licensing costs, comparability of results over time, and data confidentiality.

In mixed-mode surveys, the review finds differences in assessment outcomes between self-administered web modes and interviewer-assisted modes, especially when using cognitive measures that had originally been designed for interviewer-led administration. Since comparability between modes and over time is a priority, careful design choices are needed to reduce mode-related differences.

The review recommends prioritising cognitive measures designed for online self-completion, as this has been shown to reduce or eliminate mode effects for web-first mixed-mode surveys. In addition, choosing simpler tasks with short responses, encouraging participants to complete assessments in quiet environments, and reminding participants that honest answers are valued over perfect scores, could also contribute towards more reliable cognitive assessments in online surveys.

Finally, for studies transitioning from in-person interviewing to web-first mixed-mode approaches, conducting calibration studies, where participants complete both the legacy and new measures, is recommended. Although logistically demanding and costly, such studies may be beneficial for safeguarding data comparability across survey waves, providing evidence on mode differences for adjustment, and enabling a successful transition to web-first data collection.

1 Introduction

1.1 Overview

Cognition refers to the mental processes involved in acquiring, storing, manipulating, and retrieving information (Vasilichi, 2021). These processes support how individuals perceive and respond to their environment, guiding behaviour and decision-making. Cognitive ability plays a crucial role in shaping health, economic, educational, and social outcomes across the life course. In older adults, cognitive impairment is associated with reduced quality of life, loss of independence, increased demand for healthcare and caregiving, and often significant financial consequences (Brody *et al.*, 2019). In children and young people, cognitive development is closely linked to key developmental outcomes, including educational attainment, creativity, career success, parenting, and interpersonal relationships (Diamond, 2013).

The number of research studies directly assessing cognitive ability, both cross-sectionally and longitudinally, has grown steadily over time. Measuring cognition allows researchers to track population trends, identify disparities, detect early signs of decline, and evaluate the effectiveness of interventions. This evidence is vital for informing policy and service planning. Longitudinal surveys are particularly valuable, as they enable the study of changes in cognitive function over time. However, incorporating comprehensive cognitive assessments into large-scale surveys presents challenges. A full evaluation typically spans multiple domains, for example memory, reasoning, orientation, calculation, language, knowledge, and fluid intelligence (Ofstedal *et al.*, 2021), and relevant measures change across the life course. Many of these assessments are time-intensive, require specialised materials or controlled environments, and depend on detailed instructions or in-person administration. Consequently, interviewer-administered in-person cognitive assessment has traditionally been considered the “gold standard” in social research, as most survey-based measures are adapted from tests originally developed by psychologists for clinical settings.

With declining response rates, rising fieldwork costs and increasing online access, many surveys are adopting mixed-mode designs, often incorporating web-based options. Technological advances have facilitated the development of computerised cognitive test batteries (see reviews in Wild *et al.*, 2008; Zygouris and Tsolaki, 2014; Sternin *et al.*, 2019; Tsoy *et al.*, 2021; Zhuang *et al.*, 2021). However, evidence suggests that mode of administration and device used (when responding online) can influence results, sometimes in inconsistent ways. This complicates comparisons across participants using different modes and devices, and across time in longitudinal studies where modes may change.

Interviewer presence is a key factor contributing to mode effects in cognitive assessments and can have both positive and negative effects. First, interviewers can ensure standardised administration by adhering to test protocols, which is crucial for both data quality and measurement comparability. Interviewers can also motivate respondents and prompt quicker responses through conversational cues (de Leeuw, 2005) and ensure continued engagement with the survey. However, their presence may also introduce pressure to perform, potentially affecting outcomes negatively. Time constraints in interviewer-administered modes may also negatively impact performance. On the other hand in web surveys – i.e. in the absence of an interviewer –-respondents may use external aids such as calculators, search engines, or assistance from others (Al Baghal, 2019),. The absence of an interviewer may also reduce engagement, leading to issues such as missing data, speeding, or straight lining; however, it may also reduce performance pressures, potentially favouring higher scores from participants (Ofstedal *et al.*, 2021).

The mode of test delivery further influences assessment. In-person interviews often rely on oral delivery, sometimes supported by visual aids, while web-based assessments typically use visual interfaces with typed or clicked responses. Even subtle differences, such as using a touchscreen versus a mouse (Ofstedal *et al.*, 2021), or using a smaller screen size or device (Passell *et al.*, 2021), can affect performance.

Designing cognitive measures that provide high quality measures, are suitable for self-administration including a web context and yield comparable results across modes remains a significant challenge.

1.2 Objectives

This evidence review brings together experiences from surveys that have implemented cognitive assessments in online or mixed-mode formats, with a particular focus on web-based components.

The main objectives of the review are to:

1. Identify large-scale surveys that have incorporated web-based cognitive ability measures across different population groups, including older adults, children and young people, and the general population, globally, with a particular emphasis on the UK.
2. Describe the cognitive questionnaires and test batteries that have been adapted for web-based self-administration, highlighting their strengths, limitations, and practical considerations.
3. Characterise the data quality challenges associated with collecting cognitive data in self-administered formats, with a focus on mode effects and their implications for comparability and validity.
4. Develop recommendations and best practice guidance for adapting cognitive assessments to self-administered online formats, informed by the evidence reviewed.

Given that most cognitive assessments in large-scale surveys are interviewer-administered, the available literature on the validity and reliability of online tests, their data quality, and mode comparisons, is relatively sparse. This review draws primarily on specialised academic sources, supplemented by technical survey documentation, which offers valuable insights into practical implementation. While the scope of the review is international[1], it has a special focus on surveys from the United Kingdom. The review also contains examples drawn from the

[1] All technical reports reviewed were in English.

United States, Australia, and other European countries. Telephone-based assessments and paper-based self-administered tests are generally excluded, though they are referenced for comparative purposes, where relevant.

1.3 Scope and method of systematic searches for this evidence review

This evidence review focuses on high-quality, large-scale surveys that have implemented cognitive assessments within self-administered data collection instruments worldwide. It examines the cognitive constructs measured, the tests used, and the practical challenges of implementation. The review draws primarily on technical reports and methodological documentation, supplemented by academic literature reporting findings from cognitive assessments in population surveys.

The search covered:

- 61 longitudinal and cohort studies listed on the Cohort Network website (Society for Longitudinal and Life Course Studies, https://www.slls.org.uk/)
- 11 studies from the Gateway to Global Ageing data platform (https://g2aging.org/)
- 21 additional surveys referenced in technical reports or academic papers

In total, 93 surveys were reviewed. Of these, 75 included cognitive assessments in at least one wave and are summarised in Table 1.

Table 1. List of surveys including cognitive assessments

ID	Country	Name	Abbreviation	Type [1]	Focus [2]	Cohort Study? [3]	Self-Adm Web?
1	UK	1958 National Child Development Study	NCDS	L	LC	YES	NO
2	UK	1970 British Cohort Study	BSC70	L	LC	YES	NO
3	UK	Avon Longitudinal Study of Parents and Children	ALSPAC	L	CY	YES	NO
4	UK	Child of the New Century (Millenium Cohort)	MCS	L	LC	YES	YES
5	UK	English Longitudinal Study on Ageing	ELSA	L	A	NO	NO
6	UK	First Steps (Wirral Child Health and Development Study)	WCHDS	L	CY	YES	NO
7	UK	Growing Up in Scotland	GUS	L	CY	YES	YES
8	UK	Next Steps	Next Steps	L	LC	YES	YES
9	UK	Northern Ireland Cohort for the Longitudinal Study of Ageing	NICOLA	L	A	YES	NO
10	UK	Southampton Women's Survey	SWS	L	LC	YES	NO
11	UK	UK Biobank	Biobank	L	A	NO	YES
12	UK	Understanding Society	USS	L	GP	NO	YES
13	UK	Health and Aging in Scotland	HAGIS	L	A	YES	NO
14	UK	1946 National Birth Cohort Study	NSHD	L	LC	YES	YES
15	UK	Twins Early Development Study	TEDS	L	CY	YES	YES
16	US	Army Study to Assess Risk and Resilience in Service Members	Army STARRS	CS	Other	NO	YES
17	US	Cognition and Ageing in the USA	Cog USA	L	A	YES	NO
18	US	Early Childhood Longitudinal Study	ECLS	L	CY	YES	NO
19	US	Future of Families and Child Wellbeing Study	FFCWS	L	CY	YES	NO
20	US	Health and Retirement Study	HRS	L	A	YES	YES
21	US	High School and Beyond	HS&B	L	CY	YES	NO
22	US	National Health and Nutrition Examination Survey	NHNES	CS	GP	NO	NO
23	US	National Longitudinal Study of Adolescent to Adult Health	Add Health	L	CY	YES	NO
24	US	National Longitudinal Study of the High School Class of 1972	NLS72	L	LC	YES	NO

ID	Country	Name	Abbreviation	Type [1]	Focus [2]	Cohort Study? [3]	Self-Adm Web?
25	US	National Longitudinal Survey of Youth	NLSY	L	CY	YES	NO
26	US	National Social Life Health and Ageing Trends Study	NHATS	L	A	YES	NO
27	US	National Social Life, Health, and Aging Project	NHSAP	L	A	YES	NO
28	US	Panel Study of Income Dynamics	PSID	L	GP	NO	NO
29	US	Project Talent	Talent	L	LC	YES	NO
30	US	Understanding America Study	UAS	L	GP	NO	YES
31	US	Wisconsin Longitudinal Study	WLS	L	LC	YES	NO
32	Australia	45 and up Study	45 and Up	L	A	YES	YES
33	Australia	Generation Victoria	GENV	L	CY	YES	NO
34	Australia	Growing Up in Australia	GUA	L	CY	YES	NO
35	Australia	Household, Income, and Labour Dynamics Australia	HILDA	L	GP	NO	NO
36	Australia	The Australian Longitudinal Study of Ageing	ALSA	L	A	YES	NO
37	Australia	The Origins Project	ORIGINS	L	CY	YES	NO
38	Australia	The Raine Study	RAINE	L	LC	YES	YES
39	Brazil	Brazilian Longitudinal Study on Ageing and Well-Being	ELSI	L	A	YES	NO
40	Canada	Canadian Longitudinal Study on Aging	CLSA	L	A	YES	NO
41	Canada	Growing Up in Quebec	GIQ	L	CY	YES	NO
42	Canada	National Longitudinal Survey of Children and Youth	NLSCY	L	CY	YES	NO
43	Canada	Quebec Longitudinal Study of Child Development	QLSCD	L	CY	YES	YES
44	China	China Health and Retirement Longitudinal Study	CHARLS	L	A	YES	NO
45	Costa Rica	Costa Rican Longevity and Healthy Aging Study	CRELES	L	A	YES	NO
46	Denmark	Danish Longitudinal Survey of Youth	DLSY	L	CY	YES	NO
47	Denmark	Tracking Adolescent's Individual Lives Survey	TRAILS	L	CY	YES	NO
48	France	French Longitudinal Study of Children	ELFE	L	CY	YES	NO
49	Germany	German National Cohort Study	NAKO	L	LC	YES	NO
50	Germany	German Socioeconomic Panel	SOEP	L	GP	NO	NO
51	Germany	National Educational Panel Study	NEPS	L	CY	YES	YES
52	Hungary	Growing Up in Hungary	GUH	L	CY	YES	NO
53	India	Longitudinal Aging Study in India	LASI	L	A	YES	NO
54	Indonesia	Indonesian Family Life Survey	IFLS	L	GP	NO	NO
55	Ireland	Children's School Lives	CSL	L	CY	YES	NO
56	Ireland	Growing Up in Ireland	GUI	L	CY	YES	NO
57	Ireland	The Irish Longitudinal Study on Ageing	TILDA	L	A	NO	NO
58	Japan	Japanese Study of Aging and Retirement	JSTAR	L	A	YES	NO
59	Malaysia	Malaysia Ageing and Retirement Study	MARS	L	A	YES	NO
60	Mexico	Mexican Health and Aging Study	MHAS	L	A	YES	NO
61	Netherlands	Generation R	Generation R	L	CY	YES	YES
62	New Zealand	Growing up in New Zealand	GUNZ	L	CY	YES	NO
63	New Zealand	New Zealand Health, Work and Retirement Study	HWR	CS	A	YES	NO
64	Norway	Norwegian Mother, Father and Child Cohort Study	MoBA	L	CY	YES	NO
65	South Africa	Health and Aging in Africa: Longitudinal Studies in South Africa	HAALSI	L	A	YES	NO
66	South Korea	Korean Longitudinal Study of Ageing	KLoSA	L	A	YES	NO
67	Sweden	The Swedish Panel Study of Living Conditions of the Oldest Old	SWEOLD	L	A	YES	YES
68	Switzerland	Transitions from Education to Employment	TrEE	L	CY	YES	YES
69	Switzerland	Zurich Longitudinal Study	ZLSE	L	LC	YES	NO
70	Taiwan	Kids in Taiwan	KiT	L	CY	YES	NO
71	Thailand	Health, Aging, and Retirement in Thailand	HART	L	A	YES	NO
72	Multi-national	Survey of Health, Ageing and Retirement in Europe	SHARE	L	A	NO	NO
73	Multi-national	The Young Lives Study	YLS	L	CY	YES	YES
74	Multi-national	Survey of Adult Skills (Programme for International Assessment of Adult Competencies)	PIAAC	CS	GP	NO	NO
75	Multi-national	Programme for International Student Assessment	PISA	CS	CY	NO	NO

[1]: *Survey type: [L] = Longitudinal, [CS] = Cross-sectional*

[2]: *Focus: [LC] = Life course, [CY] = Children and youth; [A] = Ageing population, [GP] = General population*

The list contains 75 surveys from 27 countries, plus four multi-national surveys. Of these, the Survey of Health, Ageing and Retirement in Europe (SHARE) covers 28 countries in Europe plus Israel, while the Young Lives Study collects data from children living in Ethiopia, India, Peru, and Vietnam. The Survey of Adult Skills (PIAAC) and the Programme for International Student Assessment (PISA) are multinational surveys conducted by the Organisation for Economic Co-operation and Development (OCED).

Half of the national surveys were conducted in the US (16), UK (15), and Australia (7). Other countries represented include Canada, Germany, Ireland, Denmark, New Zealand, and Switzerland. Most are longitudinal studies, with only three exceptions: NHANES and Army STARRS (US), and the New Zealand HWR.

Most surveys are focused on specific population subgroups, including children and young people (29 surveys, all of which are cohort studies) and the ageing population (24, with all but two defined as cohort studies). The review also includes 12 cohort studies for the full life course, and 10 general population surveys.

We found evidence of 17 surveys conducting cognitive function assessments in self-administered online mode. Five of these surveys focus on children and young people, while three (the Health and Retirement Study in the US, the 45 and Up Study in Australia, and the SWEOLD panel in Sweden) are studies of the ageing population. The remaining surveys (9) are either life course cohort studies (e.g. the Millennium Cohort, Next Steps, and the Raine Study), or general population surveys (e.g. Understanding Society or Understanding America). Our review will mainly focus on these 17 surveys where the cognitive assessment is administered online but will also refer to other studies from the table when relevant for context and comparison.

While the primary focus of this review is on cognitive assessments within large-scale social surveys, we supplement our report with selected findings from a literature search for cognitive function assessments in smaller-scale online surveys. This search started from previous literature reviews (e.g. Wild *et al.*, 2008; Zygouris and Tsolaki, 2014; Sternin *et al.*, 2019; Tsoy *et al.*, 2021; Vasilichi, 2021; Zhuang *et al.*, 2021) from which papers reporting experiences of cognitive assessments in the context of social survey research were selected.

We review the tests conducted in these surveys separately for each population subgroup. Surveys targeting the elderly population are covered in Section 2, while surveys focused on children and youth are discussed in Section 3. General population surveys are reviewed in Section 4. Section 0 broadens the scope to include studies that explore alternative approaches to cognitive data collection. These applications offer valuable methodological insights that can inform future survey design and implementation. In Section 6, we summarise the findings from the evidence review and provide concluding remarks.

2 Surveys for the ageing population

As global life expectancy rises, the ageing population continues to grow, bringing increased attention to cognitive and physical health challenges. While some cognitive decline is a normal part of ageing, more serious conditions —such as mild cognitive impairment and age-related dementias like Alzheimer's disease—affect an estimated 20% and 7% of adults over 65 in the UK, respectively (NHS England, 2024; Alzheimer's Research UK, 2025). Declining cognitive function is often linked to functional impairment, which can reduce independence and quality of life, and increase reliance on others (Ofstedal *et al.*, 2005).

Given its impact, cognitive functioning has become a key focus in longitudinal surveys of older adults. One of the earliest and most influential studies in this area is the Health and Retirement Study (HRS), launched in 1992 in the United States. Its success inspired the development of comparable surveys worldwide (G2Aging, 2024) including the English Longitudinal Study on Ageing (ELSA), in the UK; the Survey of Health, Ageing, and Retirement in Europe (SHARE), across several European countries and Israel; and the Irish Longitudinal Study of Ageing (TILDA) in the Republic of Ireland. The National Survey of Health and Development (NSHD), the National Child Development Study (NCDS and the 1970 British Cohort Study (BCS70) are UK longitudinal birth-cohort studies following people born in 1946, 1958 and 1970 respectively. These studies administered a wide range of cognitive assessments in childhood and again in later adulthood (from Age 43 in NSHD, Age 50 in NCDS and Age 46 in BCS70). All these studies have included in-person cognitive assessments. NHSD has used web-based cognitive tests in one of their surveys (at age 77). UK Biobank, a longitudinal study focused on the health of middle-aged and older adults aged 40 to 69 when recruited between 2006 and 2010, has made use of self-administered and unsupervised online cognitive tests conducted in a lab setting.

To date, only one of these studies – HRS - has incorporated online cognitive assessments into an ongoing, longitudinal web-first mixed-mode survey protocol (see below). However, other ageing studies are considering this for future waves and the approach to the assessment of cognitive function is a key consideration which will require careful planning to ensure comparability of measurement between modes and over time as far as possible.

Section 2.1 discusses the HRS cognitive tests in detail, including their original implementation in interviewer-administered in-person and telephone interviews and then later experiences using the self-administered online mode. We discuss the NHSD in Section 2.2, and UK Biobank in Section 2.3. We briefly mention other longitudinal surveys for the ageing population that have used self-administered cognitive measures in Section 2.4.

2.1 Health and Retirement Survey (HRS)

The HRS is a biannual nationally representative survey of more than 37,000 individuals over the age of 50 in 23,000 households in the USA. It was originally established in 1992 to provide information about the economic well-being and health of this population group. The HRS is conducted by the Institute of Social Research (ISR) at the University of Michigan (Sonnega *et al.*, 2014). Baseline interviews are conducted in-person. Between 1994 and 2004, the "core" follow-up interviews were typically conducted via telephone, but in-person interviews were conducted with those aged over 80 and those who requested in-person interviews. In 2006, HRS moved to a mixed-mode design for follow-up in which half of the sample is assigned an in-person interview with physical and biological measures and a psychosocial questionnaire (the “enhanced” in-person interview), while the other half completes only the core interview, usually by telephone. The half-samples alternate waves. Cognitive assessments are included in both the enhanced in-person interviews and the regular telephone interviews.

Since 2018, online self-administration was offered as an alternative for respondents allocated to the” regular” follow-up interview. In addition, in years that do not contain a core interview, the study team fields a variety of off-year efforts, including web and mail surveys. A series of web surveys were fielded in alternate years between 2003 and 2013, containing some questions from the core interview (including the cognitive assessments), as well as a range of new topics (Ofstedal *et al.*, 2021). These latter implementations are of interest for this review, as they allowed for investigation of mode effects on performance in the cognitive assessments.

2.1.1 Cognitive measures design and rationale

HRS cognitive measures cover learning and memory, which are early indicators of cognitive decline and other abilities including reasoning, orientation, calculation, language, and knowledge where deterioration tends to occur later and can signal increased need for daily support. Cognitive measures are collected consistently across waves, allowing for comparisons over time. They cover a wide range of difficulty and are adapted from validated tools including the Mini-Mental State Examination (MMSE; Folstein *et al.*, 1975), the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), and the Woodcock-Johnson Tests of Cognitive Ability (Woodcock and Johnson, 1989). The tests used in the five most recent waves (2014 to 2022) are listed in Table 2.

Table 2. Cognitive measures collected in the Health and Retirement Survey (HRS) (2014–2022)

Domain	Test	Description
Memory	Immediate word recall Delayed 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. 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).
Attention/ Concentration	Backwards count	Respondents are asked to count backwards for 10 continuous numbers beginning with the number 20 (and from 86 in some versions).
Attention/ Concentration	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.
Orientation	Date naming	Respondents are asked to report “today’s date”.
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?”.
Language/ Naming	President/Vice-president naming	Respondents are asked to name the current President and Vice President of the United States.

Domain	Test	Description
Vocabulary	Vocabulary test (adapted from WAIS-R)	Respondents are asked to define 5 words from a closed list.
Numeracy	Questions to measure numeric ability	Three simple arithmetic problems are given to respondents to solve.
Fluid intelligence	Animal naming	Participants are asked to name as many animals as they can in 60 seconds.
Fluid intelligence	Verbal analogies (2014 and 2018)	Participants are given six verbal analogies to complete (e.g. "*mother is to daughter as father is to […]*".
Fluid intelligence	Number series test adapted from the Woodcock- Johnson (WJ-R) tests of cognitive ability (2016, 2020, and 2022)	Participants are given a series of numbers with a blank space to be completed.

*Note: These measures were originally designed to be administered by in-person interviewers and have been subsequently adapted for online self-administration.

These tests have remained largely unchanged since the survey began, providing a reliable framework for tracking cognitive changes and identifying early signs of decline in older adults. The tests were originally devised to be administered by in-person interviewers in a relatively short time (Ofstedal *et al.*, 2005)[2]. They were subsequently adapted for telephone administration and then for web administration as will be discussed in Section 2.1.2(although the backwards counting assessment is not feasible in web mode). Recent HRS web surveys have additionally included two web-specific tests – a *mouse/clicking task*, in which respondents are asked to click inside four boxes as quickly as they can and a *typing task*, in which participants are asked to type a phrase ("*The quick brown fox jumps over the lazy dog"*) as quickly as possible. These two tasks measure computer proficiency as a proxy for attention and concentration.

2.1.2 Experiences with self-administered cognitive measures

Mode effects are a significant concern when administering cognitive assessments in large-scale surveys, as they can affect comparability of scores between participants and over time when modes change. The HRS has conducted several investigations to understand differences between interviewer-administered and self-administered formats.

Ofstedal *et al.* (2021) examined experimental data from 4,223 respondents who completed the 2012 core interview (half of the respondents were randomly assigned to either a telephone interview or an enhanced in-person interview), the 2013 web interview, and the 2014 core interview (where again half of the respondents were randomly assigned to either a telephone interview or an enhanced in-person interview). survey mode. The study assessed several indicators, including missing data, completion time, score differences between modes, correlations, score trajectories over time and linear regression models of cognitive ability across tests and modes of administration, controlling for sociodemographic variables. The

[2] It must be noted that other members of the HRS International Family of Studies (https://hrs.isr.umich.edu/about/international-family-studies) use these tests for their cognitive assessments. These include ELSA and SHARE, two of the most important surveys aimed at the ageing population by scope and sample size. However, these surveys collect their data predominately using computer-assisted personal interviews, and are therefore out of the scope of this review

analysis focused on the number series test, numeracy test, Serial 7s, and verbal analogies. The findings demonstrated clear mode effects, especially between web and interviewer-administered formats, with some differences also observed between telephone and in-person interviews.

Mode appears to influence not only performance levels but also the psychometric properties (reliability and validity) of the cognitive measures. Respondents generally scored higher in web-based assessments. However, it is unclear which mode produces more valid results. Interviewer presence may cause anxiety or pressure, lowering performance. In contrast, web respondents may use external aids or take more time, potentially inflating scores. Higher rates of missing data in the web sample suggest that reduced engagement or satisficing may also play a role. Despite this, web respondents tended to spend more time on tasks and achieved higher scores than those interviewed in-person.

Domingue *et al.* (2023) conducted a related experiment during the 2018 HRS wave. Respondents who had previously completed in-person interviews in 2016 were randomly assigned to either phone or web modes in 2018. The study found consistently higher cognitive scores among those assigned to the web mode across all items. Although the authors did not test specific causes, they suggest that differences in how questions are presented may contribute. Web respondents can review questions and response options visually, while telephone respondents typically hear them only once, which may affect comprehension and performance.

2.2 1946 National Birth Cohort (National Survey of Health and Development)

The 1946 National Birth Cohort survey, also known as the National Survey of Health and Development (NSHD), was the first ever British birth cohort study. The NSHD is a representative sample of over 5,000 males and females who were born in England, Scotland, and Wales in one week in March 1946 (Wadsworth *et al.*, 2006). The study started with information collected by health visitors on all births during that week. Subsequently, the study has collected sociodemographic, medical, cognitive, and psychological functioning data through interviews and examinations in 27 waves, as well as smaller sub-study collections. As of 2024, there are approximately 2,700 participants in active follow-up (Cai *et al.*, 2024). The current aim of the NSHD study is to explore long-term ageing and how it is affected by factors across the life course. Cognitive tests have been a key feature of NSHD with assessments primarily being conducted via in-person interviews.

In 2023, NSHD invited participants with internet access and an email address to complete an online battery of cognitive assessments using Cognitron, an online platform designed for remote cognitive testing[3] (Cai *et al.*, 2024; Giunchiglia *et al.*, 2025). The study aimed to assess the uptake, adherence, and usability of online cognitive assessments in a sample of older adults who were active members of the NHSD sample at the time. Those without internet access or an email address were not invited to participate in this study.

[3] More info can be found in https://www.cognitron.co.uk/.

Cognitron is a platform that hosts over 100 optimised cognitive tasks, designed to be sensitive, domain-specific, and validated across both general and clinical populations (Hampshire *et al.*, 2024).The tasks are selected to assess a range of cognitive functions, including reaction time, motor control, memory, attention, reasoning, and executive function. Table 3 lists the specific tasks included in the NSHD study, as reported by Del Giovane *et al.* (2025).

The battery covered a range of cognitive domains, including working memory, measured through immediate and delayed object recognition; language, assessed via word definitions and verbal reasoning; and attention and concentration, evaluated using spatial span tasks. Computerised assessments like Cognitron are particularly well suited to measuring processing speed, executive function, and visuospatial abilities, which can be difficult to measure in other contexts.

Participants accessed the tasks through a web browser on any smartphone, tablet, or PC/laptop computer. The battery was presented in a fixed order, with general instructions at the start and specific instructions before each task, followed by short practice trials.

1,753 members of the NSHD cohort (all aged 77) were invited to participate, with 990 (56.4%) providing consent, and 813 attempted the battery (46.4% of those invited). Of the 813 participants who began the battery, 88.8% completed all 13 tasks, with a median completion time of 41 minutes.

The study reported high levels of consent, participation, and completion. These outcomes were associated with sociodemographic and health-related factors. Higher education was linked to greater likelihood of consenting, better understanding of the study, and increased confidence in task performance.

Cai *et al.* (2024) noted that some participants in the top 10 percent for both response time and accuracy were flagged as possible cheaters, as they spent excessive time clicking outside the task browser during the assessment, indicating a possible use of external websites or help. The study also gathered qualitative feedback. Participants generally requested clearer instructions and more user-friendly interfaces. Notably, most participants who began the online assessment completed it. The greatest barrier appeared at the recruitment stage. Cai *et al.* (2024) recommend streamlining the transition from recruitment to task completion to reduce dropout caused by switching between platforms or devices.

Table 3. Cognitive measures collected via Cognitron in the 1946 National Birth Cohort study (2023)-

Test name	Domain	Time (min)	Summary
Objective immediate and delayed recognition	Working memory	2–3	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.
Motor control	Processing speed	3	Participants are shown a red target appearing at different locations of the screen and asked to tap on it as quickly as possible.
Choice reaction time	Processing speed	2	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.
Blocks	Visuospatial abilities	3	Participants are asked to remove blocks of different colours and shapes from one array to match the target array.
Digit span	Executive function	4	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.
Spatial span	Visuospatial abilities, Attention	2	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.
Stroop	Executive function	5	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).
2-D manipulations	Visuospatial abilities	2	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.
Word definitions	Language	3	Participants are shown a word and 4 possible definitions and asked to tap on the correct definition within a designated amount of time.
Verbal reasoning	Language	3	Participants are shown different combinations of geometric shapes and asked to indicate whether the statement describing the shapes is true or false.
Spotter	Processing speed	5	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.
Forager	Processing speed	3	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).

*Note: These measures were designed for online self-administration.

2.3 Biobank (UK)

UK Biobank is a large prospective cohort study investigating the health of middle-aged and older adults in the UK (Fawns-Ritchie and Deary, 2020). Between 2006 and 2010, it recruited 500,000 volunteers aged 40–69. Although not representative of the UK population, it is relevant to this review due to its use of web-based cognitive assessments.

Initial cognitive testing included brief tasks such as the pairs memory test (visual memory) and reaction time test (processing speed). Subsamples also completed tests of working memory, prospective memory, and fluid intelligence, similar to those used in the HRS.

Since 2014, UK Biobank has conducted an imaging study involving over 100,000 participants, who repeat the baseline assessment and undergo brain and body scans. These are completed unsupervised via a fully automated touchscreen interface in a lab setting. The battery includes tests of processing speed (reaction time, symbol digit substitution), memory (pairs matching, prospective memory, numeric memory, paired associate learning), and executive function (trail making, tower rearranging). It also assesses crystallised ability[4] and non-verbal reasoning. These domains overlap with those in the Cognitron battery (used in the 1946 National Birth Cohort) and the CogState battery (used in the Raine Study, see Section 3.3 for details)[5]. Additional measures include a self-rated memory questionnaire and the Mini Addenbrooke's Cognitive Examination (Hsieh *et al.*, 2014), which screens for early dementia using tasks like date recall, address memorisation, animal naming, and clock drawing. These items closely resemble those used in HRS cognitive assessments. The tests used in the UK Biobank survey are listed in Table 4.

Fawns-Ritchie and Deary (2020) found strong concurrent validity between the UK Biobank cognitive battery and established standard tests of cognitive ability. Most tests showed moderate-to-good test-retest reliability, a measure of consistency that evaluates whether similar results are obtained when the test is administered to the same individuals on different occasions. Notably, the general cognitive ability score derived from the UK Biobank battery correlated at 83% with a comparable score from reference tests. However, the four-week test-retest reliability was only moderate, with a correlation of 55%. While these findings support the potential of self-administered computerised cognitive assessments, they are based on a small sample of 160 volunteers who were already inclined to participate in research. This limits the generalisability of the results to the broader population.

---

[4] *Crystallised ability* refers to the application of learned procedures and knowledge, which depends on experience and education. It is used in opposition to *fluid ability*, which refers to the ability to solve novel reasoning problems, and depends on skills such as comprehension, problem-solving, and learning (Unsworth et al., 2014).

[5] Fawns-Ritchy and Deary (2020) state that, beyond a brief description of each test and some basic statistics on the Biobank website, "*there is little other information regarding why these specific tests were chosen or how these tests were developed*" (p. 2). Fawns-Ritchy and Deary (2020) are the first to conduct a formal analysis on concurrent validity and test-retest reliability for the Biobank test battery.

Table 4. Cognitive measures collected in the Biobank study (2014 onwards)

Test	Domain	Description
Pairs matching test	Visual declarative memory	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.
Reaction time test	Processing speed	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.
Prospective memory test	Prospective memory	Early in the touchscreen cognitive section, the participant is 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.*"
Fluid intelligence test	Verbal and numerical reasoning	Participants have 2 minutes to complete as many questions as possible from the test.
Numeric memory test	Working memory	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).
Trail making test parts A and B	Executive function	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.
Symbol digit substitution test	Processing speed	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.
Picture vocabulary	Crystallised ability	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.
Paired associate learning test	Verbal declarative memory	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.
Tower rearranging test	Executive function	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.
Matrix pattern completion	Non-verbal reasoning	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.

*Note: These measures were designed for online self-administration.

2.4 Other surveys

Aside from the previously reviewed surveys (namely the HRS, 1946 National Birth Cohort Study, and Biobank), our review of technical reports identified only a limited number of examples of cognitive assessments being included in self-administered surveys (without an interviewer present) targeting the ageing population. For example, while the Swedish Panel Study of Living Conditions of the Oldest Old (SWEOLD) is mainly conducted via telephone interviews, it also offers self-administered paper-based questionnaires as an alternative for participants (Lennartsson *et al.*, 2014; Ramos-Serrano and Fors, 2024). The survey includes a reduced version of the MMSE consisting of immediate and delayed recall (three items), orientation (date and country), and the Serial 7's test. However, the survey has not yet moved to online self-administration. Similarly, the 45 and Up Study in Australia (Bleicher *et al.*, 2023) is self-administered via paper questionnaires, and Wave 4 (in the field between 2023 and 2025) collects information about "*experience of memory loss and diagnosis of cognitive impairment, dementia or Alzheimer's*"[6]. A randomised controlled study nested within the 45 and Up Study, called Maintain Your Brain, aimed at reducing cognitive decline with ageing using an online package of interventions administered intensively for 12 months, followed by monthly boosters for two months. The trial used the CogState battery of cognitive assessments (introduced in detail in section 3.3). No results are available for either of those data collection processes at the time of writing this review

3 Surveys of children and young people

Many longitudinal studies seek to measure cognition during childhood because it provides a foundation for understanding how cognitive abilities evolve over time and how they relate to later life outcomes.

Numerous studies have administered cognitive assessments with very young children. For example, the Avon Longitudinal Study of Parents and Young Children (ALSPAC; Golding *et al.*, 2001) administered measures of visual attention and visual recognition at 4 months and the German National Education Panel Survey (NEPS) included habituation-dishabituation tasks in follow-ups at 7 and 17 months. These tasks assess cognitive abilities by observing attention to repeated stimuli. Perhaps unsurprisingly, due to the significant practical difficulties that would be entailed, we find no evidence however of any studies attempting to measure early years cognition in an online context, thereby out of scope for this review.

This section covers longitudinal, and cohort surveys based on children and the youth, which often measure cognitive abilities as part of their interviews. Cognitive responses for these groups are mostly collected directly from the participants, and the test batteries are chosen to accurately measure cognitive development at each age level. While most studies listed in Table 1 are interviewer-based, this chapter focuses on experiences in online self-administered cognitive function testing in surveys including The Millennium Cohort Study and Next Steps in the UK (Section 3.1), the National Educational Panel Study in Germany (Section 3.2), the Raine Study in Australia (Section 3.3), and the Twins Early Development Study (TEDS) (Section 3.4). Other longitudinal and cohort surveys for this age group are discussed in Section 3.5.

This review excluded cognitive tests included in international educational assessments such as PISA, TIMSS, and PIRLS. Given the classroom environment testing, the cognitive assessment

[6] This is reported on the website for Wave 4: https://www.saxinstitute.org.au/solutions/45-and-up-study/use-the-45-and-up-study-2/45-and-up-study-wave-4/

tasks in these assessments are not conducted online and are therefore out of scope for this review.

3.1 The Millennium Cohort Study and Next Steps (UK)

The Millenium Cohort Study (MCS) follows the lives of around 19,000 young people born across England, Scotland, Wales, and Northern Ireland in 2000–2002. Follow-ups conducted between Ages 3 and 17 were all conducted in-person and all included cognitive assessments. The latest sweep, the Age 23 Survey which took place between 2022 and 2024, was however conducted as an online-first mixed mode survey with non-respondents invited to take part in an in-person interview.

Next Steps, previously known as the Longitudinal Study of Young People in England (LSYPE), follows the lives of around 16,000 people in England born in 1989– 1990 (Wu *et al.*, 2024). The study began in 2004, the most recent sweep (Age 32) took place between 2022 and 2023 using an online-first mixed mode approach with web non-respondents followed up in-person, with telephone and video modes also available. The Age 32 sweep was the first sweep in Next Steps to include a cognitive assessment.

The MCS Age 23 Survey and the Next Steps Age 32 Survey both used the same Backwards Digit Span (BDS) test. The BDS task measures working memory and attention and was developed by not-for-profit organisation TestMyBrain (Singh *et al.*, 2021). In this test, participants are displayed sequences of digits of increasing length, which they are then asked to recall in reverse order. After the final digit is displayed, participants are asked to type in their response. The sequences start at two digits and increase to a maximum of 11, with the task discontinued if two trials at the same length are failed. The average completion time is 3.5 minutes (Singh *et al.*, 2021).

The MCS Age 23 Survey also used the Stroop test – though this was not included in the Next Steps Age 32 Survey. The Stroop test is a widely used cognitive test that measures attention, processing speed, and executive control. Participants are presented with a series of words describing colours, but the words are printed in different colours, creating a conflict between reading the word and identifying the colour of the text. Participants are asked to record the colour of the text rather than the word. For example, if the word ‘blue’ is written in red text the correct answer would be ‘red’. Delays in response time or errors reflect the individual's ability to inhibit automatic responses and manage cognitive interference.

The BDS (in both surveys) and the Stroop test (in MCS) were administered as part of the web survey for online participants. During in-person interviews the same assessments) were completed via self-completion on the interviewer’s tablet.

Online participants were re-directed from the web survey to a website hosted by TestMyBrain to complete the BDS, whereas in-person participants completed the task using software installed on the interviewer’s tablet because during interviews tablets were not connected to the internet. The respondent interface was however identical in both modes. For MCS, the Stroop task was however programmed into the web survey by the fieldwork agency.

The Centre for Longitudinal Studies, UCL which runs both MCS and Next Steps recently conducted a methodological project which sought to understand the extent to which mode effects might impact some of the key measures collected in the most recent sweeps of the two studies – including the BDS. A convenience sample of participants aged 20-40 living in England ($n = 1800$) was recruited using quotas and asked to complete two surveys, both of which included the BDS, approximately two weeks apart. Participants were randomly allocated to web, video or in-person modes at both time points creating nine mode sequences. The BDS was completed via self-completion in all modes. This experimental setting will allow for robust assessment of how performance in the assessment is affected by mode. Encouragingly, the results, which are in preparation for publication show that mode effects are minimal (Tsigaridis *et al.*, 2025).

3.2 National Educational Panel Study (Germany)

The German National Educational Panel Study (NEPS) is a longitudinal multi-cohort study launched in 2009 that tracks the development of domain-specific competencies from birth to adulthood in a nationally representative sample. NEPS includes a wide range of cognitive measures for children and young people, such as reading and mathematical competence, scientific literacy, ICT literacy, and reading skills, both in German and in English (as a foreign language).

NEPS has explored the feasibility of self-administered cognitive testing. In Wave 5 of Starting Cohort 5, Zinn *et al.* (2021) conducted an experiment with university students randomly assigned to one of three groups: supervised paper-based testing, supervised computer-based testing, and unsupervised web-based testing. Students originally assigned to the supervised versions of the tests but who refused participation were subsequently invited to complete the web-based test. The scientific literacy test measured understanding of basic scientific concepts and processes.

Contrary to previous research, students who were randomly assigned to the modes showed notably higher response rates in unstandardised and unsupervised web-based assessments (54.2%) as compared to standardised and supervised assessments (25.6% for paper-based and 18.2% for computer-based tests). The analysis also shows that the unsupervised assignment does not introduce different selection biases as compared to the supervised versions of the test. Measurement properties were largely consistent across modes, though reliability was slightly lower in the web-based format. A small systematic bias was observed, with paper-based assessments yielding slightly higher scores (less than 2.5% of the total score), particularly among lower-ability participants. Overall, the findings support the feasibility of unsupervised web-based cognitive assessments.

In Wave 12, Starting Cohort 5 participants completed self-administered web-based tests in mathematics, German reading, and English competence. Half of the respondents completed the test in a invigilated setting at their private homes, while the remaining participants worked in a non-invigilated setting. Gnambs (2019) evaluated the English competence module and found it had acceptable psychometric properties, supporting its use for reliable competence estimation.

These findings suggest that while minor mode effects exist, web-based cognitive assessments can be a viable alternative to supervised formats, especially in large-scale longitudinal studies where flexibility and cost-efficiency are important.

3.3 The Raine Study (Australia)

The Western Australian Pregnancy Cohort (Raine Study) is a multigenerational longitudinal study that began in Perth in 1989 and now spans four generations. It has used CogState, a widely adopted computerised cognitive test battery, to assess cognitive function, including working memory, executive function, and attention/concentration. CogState is designed to be repeatable, sensitive, and efficient, with a format that is culturally neutral and language insensitive (Allen *et al.*, 2012). Research has shown strong correlations between CogState scores and traditional neuropsychological tests, as well as high sensitivity to cognitive impairment across various conditions (e.g. Maruff *et al.*, 2009; Lupu *et al.*, 2021).

The 17- and 22-year-old follow-ups of the Raine Study included CogState tasks assessing memory and attention, as well as more complex domains such as visuospatial abilities, processing speed, and executive function. The tests were self-administered via a web-based survey. Table 5 provides a full list of the assessments conducted.

Table 5. Cognitive measures collected via CogState in the Raine Study 17- and 22-year follow-up surveys (2013-2014 and 2018-2019)

Test name	**Domain**	**Time (min)**	**Summary**
Continuous Paired Associate Learning Test (CPAL)	Visuospatial abilities, Working memory	7	Participants must learn and remember pictures hidden beneath different locations on the screen.
Detection Test (DET)	Processing speed	3	Measures processing speed using a simple reaction time paradigm ("has the card turned over"?)
Groton Maze Learning Test (GMLT)	Executive function, Administration	7	A 10 x 10 grid of tiles is presented on the screen, with a 28-step pathway hidden among these tiles. The 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.
Identification Test (IDT)	Attention/ Concentration	3	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.
One Card Learning Test (OCL)	Working memory	6	A playing card is presented face up in the centre of the screen and the participant must decide whether they have seen it before in the test.
One Back Test (OBT, *22-year follow up only*)	Working memory	4	A playing card is presented face up in the centre of the screen. The participant must decide whether the card is the same as the previous card.
Set-Shifting Test (SET, *22-year follow up only*)	Executive function	7	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

Test name	Domain	Time (min)	Summary
			"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.

*Note: These measures were designed for online self-administration.

While the CogState battery was developed for self-administered use on computers, studies support its validity and acceptability on other devices, including iPads (Mielke *et al.*, 2015). Although performance differences across devices were relatively small, the mode of input—keyboard, mouse, finger, or stylus—affected speed and accuracy. Participants were faster and more accurate using a keyboard or mouse on a PC compared to finger touch on an iPad, though they preferred the iPad and believed they performed better on it (Mielke *et al.*, 2015; Stricker *et al.*, 2019).

More recently, CogState has been tested on smartphones (Edgar, 2023; Cummins *et al.*, 2025), with high levels of reported usability and acceptability. Performance was slower on smartphones than on computers, and while over 85% of participants found the text and button size appropriate, there was no strong preference for smartphones. Issues such as fatigue, distraction, and input method challenges were noted across both platforms.

3.4 Twins Early Development Study (TEDS)

The Twins Early Development Study (TEDS) is the first large-scale population-based twin study in the UK. It focuses on the early development of the three most common psychological problems in childhood: communication disorders, mild mental impairment, and behaviour problems. More than 15,000 pairs of twins have been enrolled in TEDS, and they were identified from birth records of twins born in England and Wales in 1994, 1995, and 1996. The participating families are representative of the UK (Trouton *et al.*, 2002). Since first contact, data have been collected at 2, 3, 4, 7, 8, 9, 10, 12, 14, 16, 18, 21, and, most recently, 26 years (Lockhart *et al.*, 2023). Cognitive and language measures were administered to twins at several of these waves, including:

- parent-administered tests at ages 2, 3, and 4
- in-home, interviewer-administered tests at age 4
- phone tests at age 7
- questionnaire self-administered tests at age 9
- web self-administered tests at ages 10, 12, 14, 16, 18, 21, and 26

As cognitive ability is one of the focus of TEDS, the study has used a wide range of measures. Table 6 lists the self-administered online test batteries used in the context of TEDS. The table indicates the age group to which each battery was applied. All these tests were specifically designed to be self-administered via web.

Table 6. Cognitive measures collected in TEDS (2003-2022)

Ages	Test name	Domain	Description
10, 12, 14, 16	Ravens Progressive Matrices	Non-verbal reasoning	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.
10,12	WISC: Picture completion	Non-verbal reasoning	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.
10, 12, 14	WISC: Vocabulary	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.
10, 12, 14	WISC: General knowledge	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,
10, 12	Author recognition	General knowledge	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.
12,16	Woodcock-Johnson III Reading Fluency	Reading ability	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.
12	GOAL Formative Assessment in Literacy (Key Stage 3)	Reading ability	Multiple choice of reading comprehension ability.
10, 12, 16	Number games	Mathematics	The test included 3 sub-tests: understanding numbers (33 items), non-numerical processes (25 items), and computation and knowledge (37 items).
12, 16	Test of Language Competence (Expanded Edition). Semantics: The Figurative Language subset	Language competence	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.
12	Test of Language Competence (Expanded Edition). Pragmatics: The Making Inference subset	Language competence	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.
12	Listening Grammar Subtest of the TOAL-3.	Language competence	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
16, 21	Mill Hill Vocabulary test	Vocabulary	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.
16	Passages comprehension	Reading ability	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

Ages	Test name	Domain	Description
12	Hidden shapes test	Spatial reasoning, Figure recognition	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.
12	Jigsaws tests	Spatial reasoning, Figure recognition	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.
12	Eyes test	Social sensitivity, Mind ability	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.
14	Science test	Science, General knowledge	A test of Scientific Enquiry skills, based on the UK National Curriculum. The test included 39 items.
18	Bricks test	Mental rotation, Visualisation	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.
18	Kings Challenge test	Spatial abilities	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.
18, 26	Navigation web study	Spatial abilities	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.
26	TestMyBrain Vocabulary test	Vocabulary	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.
26	TestMyBrain Digit Symbol Matching test	Processing speed	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.
26	TestMyBrain Verbal Paired Associates Memory test	Verbal declarative memory	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
26	TestMyBrain Forward and Backward Digit Span test	Executive function	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).

*Note: These measures were designed for online self-administration.

The tests applied at age 26 are similar to test batteries used in other contexts. For example, the Digit Span test is used in Cognitron (Table 3), while the Verbal Pair Associates memory test and the Symbol Matching test are used in Biobank (Table 4). In contrast, the tests used for younger participants are unique to TEDS. All these tests were originally designed for online self-administration, and their results have enabled the study of how genetics and environment shape the development of cognitive abilities (Rimfeld *et al.*, 2019; von Stumm *et al.*, 2023; Knyspel and Plomin, 2024).

In addition, a sample of 4,751 21-year-old twins from TEDS completed Pathfinder. This gamified, 15-minute test measures general cognitive ability, the underlying dimension connecting various specific cognitive functions. Pathfinder is comprised of two core blocks assessing verbal and non-verbal abilities through 20 items each. These items span five cognitive tests (visual puzzles, matrix reasoning, verbal analogies, vocabulary, and missing letters) embedded within a gamified storyline (Malanchini *et al.*, 2021). The measures derived from Pathfinder demonstrated reliable verbal and non-verbal scores, which correlated substantially with standard cognitive measures collected at earlier ages in the study. These early findings suggest that Pathfinder offers a compelling and engaging alternative for measuring cognitive abilities, particularly suitable for children and younger participants[7].

3.5 Other longitudinal and cohort surveys

Generation R (Kooijman *et al.*, 2017) is a population-based prospective cohort study from foetal life until adulthood in a multi-ethnic population in the Netherlands. The study is designed to identify early environmental and genetic causes and causal pathways leading to normal and abnormal growth, development and health from foetal life, childhood and young adulthood. Besides cognitive questionnaires for parents in the early infancy interviews, their "Focus at 13" survey included assessments of cognitive function of teenagers using a computer game environment. The results are not published and consequently not included in this review.

Transitions from Education to Employment (TREE) (Gomensoro and Meyer, 2017) is a multi-cohort, multi-disciplinary panel study of compulsory school leavers in Switzerland, focusing on educational and occupational trajectories. The survey includes an adaptation of the cognitive ability test developed by Heller and Perleth (2000) to assess the cognitive capabilities of students from grades between 4 and 12 (ages approximately between 10 and 18 years old). Specifically, TREE adopted the figural or non-verbal subtest N2, which measures reasoning using pairs of figures or drawings that share a logical relationship. The respondents' task consists of determining the relationship (analogy) between the figures. This non-verbal test was chosen to ensure comparability across Switzerland's various language regions. Another advantage is the relatively short administration time. The N2 test was adapted from paper-and-pencil to web-based self-administration for the second tree cohort (TREE2). Despite the complex introduction translated into two additional languages (Italian and French), the test scores aligned closely with those from other studies. Some disparities were nevertheless observed, particularly among students attending programmes with low academic requirements and those from Italian-speaking Switzerland (Krebs-Oesch *et al.*, 2023).

---

[7] See Section 4.2 for a discussion on SeaHeroQuest, another gamified cognitive assessment application, implemented in the context of Understanding Society's Innovation Panel.

The TestMyBrain tests, used in the Millenium Cohort Study, Next Steps (Section 3.1), and TEDS (Section 3.4) are hosted on the TestMyBrain.org platform. The tests have been widely assessed, demonstrating high validity and consistency with tests taken in lab or clinical settings (Singh *et al.*, 2021). The tests have also been used to investigate differences in cognitive scores across devices. An analysis of cognitive test scores from about 60,000 voluntary participants (2014–2019) demonstrated that users of mobile devices (particularly those using Android smartphones) performed significantly slower on tests of reaction time than laptop and desktop users. Mobile users also tended to score lower on vocabulary accuracy (Passell *et al.*, 2021). These differences remained significant even after controlling for sociodemographic characteristics and may be related to aspects such as operation system, input type (touchscreen vs. mouse), and screen size. The results suggest that device type should be accounted for when analysing self-administered online cognitive data.

4 General population (cross-sectional and longitudinal) surveys

As discussed in Al Baghal (2019), cognitive ability is rarely measured directly in large-scale general population surveys, which often rely on proxies such as age or education. Nevertheless, there are numerous examples of surveys, mostly longitudinal, that do include cognitive assessments in at least some waves. As shown in Table 1, examples include the Household, Income, and Labour Dynamics in Australia (HILDA) survey, the German Socio-Economic Panel (SOEP), the Panel Study of Income Dynamics (PSID), the Indonesian Family Life Survey (IFLS), the National Health and Nutrition Examination Survey (US), and Understanding Society (UK). Most of these are interviewer-administered in-person survey (either computer-assisted or paper-based) and fall outside the scope of this review. PSID, Understanding Society, Understanding America, and the Real Time Assessment of Community Transmission (REACT) study have administered cognitive assessments online as described in the sections that follow.

4.1 Panel Study of Income Dynamics (US)

The Panel Study of Income Dynamics (PSID), launched in 1968 by the University of Michigan, is the world's longest-running household longitudinal panel survey. It tracks individuals and families over time, collecting data on employment, income, health, education, and family structure. While the core survey is typically administered in-person or by telephone, recent supplemental modules have explored mixed-mode approaches, including web-based surveys.

One such module is the 2016 Well Being and Daily Life Supplement (PSID-WB), which focused on wellbeing, personality traits, and everyday skills (Freedman, 2017). The everyday skills section included four cognitive assessments: verbal reasoning, health literacy, quantitative reasoning, and financial literacy.

- *Verbal reasoning*. A series of sentence completion questions, drawn from the 1972 version of the PSID. Participants are required to select the word that makes the most sensible complete sentence.

- *Health literacy*. This section measures how well respondents understand health care materials. The items were drawn from a scale known as the Test of Functional Health Literacy in Adults (Parker *et al.*, 1995).
- *Quantitative reasoning*. This section uses the number series test from the Woodcock-Johnson (WJ-R) tests of cognitive ability, also used in the HRS (see section 2.1). Importantly, this section differs across modes of survey completion. Web respondents receive two blocks of three items each, and the difficulty level of the second block depends on the score on the first block. Paper respondents receive seven items in all, three from the first block and then one item from each potential follow-up block.
- *Financial literacy*. This section uses items that measure the use of math skills in daily life. The items were previously used in HRS and ELSA.

The objective of this supplement is to explore how economic, social, and health outcomes interact with overall well-being across the life course, particularly within families (Johnson *et al.*, 2019). It used a mixed-mode, self-administered design, initially offered online, with paper questionnaires introduced later. These data have been widely used in the literature to examine links between cognition and outcomes such as well-being, employment, and financial status (e.g. Kobayashi and Feldman, 2019; Chan *et al.*, 2024). These substantive papers do not mention any reports of potential mode effects on cognitive measures.

4.2 Understanding Society (UK)

Understanding Society is the UK's main large-scale household longitudinal survey, led by the Institute for Social and Economic Research at the University of Essex. Launched in 2009, it covers around 40,000 households, with fourteen waves of data collected to date (Kantar Public and National Centre for Social Research (NatCen), 2022). The study also includes the Innovation Panel, a separate sample of approximately 1,500 households used to test new survey methods and research areas.

Cognitive assessments have been included in both the core survey and the Innovation Panel. In Wave 3 of the main survey (2011–2012), a cognitive module assessed memory, concentration, numeracy, and literacy using tasks such as word recall, subtraction, number sequences, verbal fluency, and numerical ability. The Innovation Panel Wave 3 added three further tasks: the F-A-S verbal fluency test (Patterson, 2011), a prospective memory test, and the Serial 7s test—all administered by interviewers (McFall, 2013). In Wave 10 (2019–2020), the youth questionnaire included the Raven's Standard Progressive Matrices, a paper-based test measuring abstract reasoning through pattern recognition.

Computerised cognitive testing was introduced in Wave 7 of the Innovation Panel (Al Baghal, 2019). Two-thirds of households were assigned to a mixed-mode design (web followed by in-person), while the rest completed standard in-person interviews. In both modes, cognitive tests were self-administered using the same visual interface. In in-person settings, interviewers turned the screen for respondents to complete the tasks independently. The module included number series, verbal analogies, and numeracy questions, adapted from the HRS (see Section 2.1).

Al Baghal (2019) found significant mode effects: web respondents consistently performed better than those interviewed in person across all cognitive measures. These differences persisted even after controlling for sociodemographic factors and prior cognitive scores, raising concerns about comparability. Mode-related measurement differences may distort indicators of cognitive ability across and within respondents and over time.

Respondents in Wave 16 of the Innovation Panel were invited to download and play a cognitive testing app called Sea Hero Quest (Coutrot *et al.*, 2018). This smartphone and tablet-based video game was designed for Alzheimer's Research UK to help advance the understanding of spatial navigation, which is one of the first cognitive skills affected by dementia (Burton *et al.*, 2024). The game measures way finding and path integration, two key dimensions of spatial ability, across a series of progressively challenging levels (Coutrot *et al.*, 2024). After briefly viewing a map, players must navigate a boat from a starting point to a flag's location shown on the map, while the app records their completion time. The game consists of seventeen levels set in different environments. It yields a wayfinding score that serves as an indicator of cognitive function and has been shown to correlate with variables such as sex, age, and education level (Spiers *et al.*, 2023). The app also collects metadata, including time spent on each task and the number of levels completed.

Of the 2,694 panel participants, 47.3% downloaded and started using the app. As part of an experimental incentive, payment amounts were conditional on playing the game. Uptake was 42.9% among those offered a £10 incentive and 51.4% among those offered £30. The higher incentive was also more effective at encouraging players to finish the game: 57.4% of the £30 group completed 100% of the game, compared to 47.4% of the £10 group (Burton *et al.*, 2024). The results of this study can provide a foundation for developing future cognitive measures that employ alternative data collection methods rather than conventional tests. More details about these applications are included in Section 5.2.

4.3 Understanding America (US)

The Understanding America Study (UAS) is a probability-based internet panel managed by the Center for Economic and Social Research at the University of Southern California (Kapteyn *et al.*, 2024). It includes over 15,000 participants recruited via address-based sampling. As of August 2024, more than 640 distinct surveys have been conducted. The UAS core surveys incorporate the full HRS instrument, along with modules on financial literacy and 11 cognitive measures.

Cognitive tests in the UAS include several adapted from the HRS, such as verbal analogies, number series, word recall, serial 7s, and orientation and naming tasks (e.g., date, object, and President/Vice-President naming). In addition to these traditional measures, UAS has implemented newer assessments like the stop-and-go switching test and the figure identification test.

The stop-and-go switching test, adapted from the Brief Test of Adult Cognition by Telephone (BTACT; Lachman *et al.*, 2013), measures attention and executive function, including reaction time, task switching, and inhibitory control. In the UAS, it was programmed for web self-administration and adapted for both keyboard and touchscreen devices. Respondents press “S” for “stop” or “G” for “go” based on the word displayed. Initial results showed slower response times on touchscreens compared to keyboards, prompting refinements such as practice trials, simplified instructions, and interface adjustments. Subsequent experiments found no significant device-related differences (Liu *et al.*, 2022).

The figure identification test, originally a paper-based task, measures perceptual speed. Participants identify which of five figures matches a target image. The test was adapted for web use and programmed for devices with keyboards (such as desktop and laptop computers) and touchscreens (tablets and mobile phones). A randomised experiment had participants complete all 60 items twice—once on a keyboard and once on a touchscreen. While response times did not differ significantly by device, a notable training effect was observed: participants performed faster on their second attempt. However, accuracy (number of correct responses) remained consistent across modes (Kapteyn *et al.*, 2021).

4.4 Real Time Assessment of Community Transmission (REACT UK)

The Real Time Assessment of Community Transmission (REACT) cohort study in England tracked the prevalence of COVID-19 in England from May 2020 to March 2022 using data from a random community sample of adults aged 18 or over (Hampshire *et al.*, 2024). Between August and December 2022, a sub sample of study participants was selected for a follow-up survey and a cognitive assessment. The assessment featured eight tasks from the Cognitron battery, presented in a fixed order: immediate memory, 2D mental manipulation, spatial working memory, spatial planning, verbal analogical reasoning, word definitions, information sampling, and delayed memory. Cognitive ability was assessed using both test scores and secondary measures, such as response times and error types.

800,000 REACT study participants were invited to take part in a follow-up web survey focused on health and well-being, of which 276,840 respondents (34.6%) did so. At the end of the web survey, participants were asked if they were willing to respond to do the cognitive assessment and, if they agreed, they were given a link to the website to complete them. 141,583 (51.1%) questionnaire respondents started the cognitive battery (i.e., completed at least one task), and 112,964 (79.8%) of those who started, completed all eight tasks. The report does not specify uptake or completion rates for specific sociodemographic groups. Participants were allowed to use any personal device (desktop or laptop computer, tablet, or mobile phone) to complete the assessment; however, the published reports do not examine score differences across devices. This implementation demonstrates the feasibility of online cognitive testing for very large samples.

5 Cognitive assessments in other surveys

In this section, we review applications conducted outside the scope of large-scale social surveys. We select these applications because their methods and results highlight methodological and practical implications for implementation. Computer-based cognitive assessments administered in research facilities are discussed in Section 5.1. Alternative modes of cognitive data collection are introduced in Section 5.2, while Section 5.3 focuses on studies that explore how survey response behaviour can serve as a proxy for cognitive function.

5.1 Studies in experimental settings

Cognitive function assessments are widely used in medical and psychological research, particularly in fields such as neurology, psychiatry, geriatrics, and psychology. While most of these studies are not population-representative they can still offer valuable insights into the feasibility of self-administered cognitive testing and the impact of mode and device on test performance. This section reviews two such studies.

The Army STARRS (Ursano *et al.*, 2014) is a large-scale research project investigating risk and resilience factors for suicide and mental health among U.S. Army soldiers. One component, the New Soldier Study, assessed cognitive and emotion-processing domains using laptop-based tests administered in group settings at research facilities. Five cognitive measures and two emotion-processing tasks were used (see Table 7). Moore *et al.* (2019) report successful administration to over 50,000 soldiers with minimal complications. Despite its cross-sectional design and limited generalisability, the battery demonstrated strong psychometric validity and may be adaptable for online self-administration.

Gooch (2015) conducted a large-scale randomised experiment comparing interviewer-led and self-administered computer-based surveys. Participants completed the Wordsum test, a 10-item verbal intelligence measure used in the U.S. General Social Survey. In-person interviews were conducted in mock living rooms in a research facility, while self-completion surveys took place in private office-like rooms. Gooch (2015) found mode effects linked to question difficulty: easier items were answered more accurately in in-person settings, while harder items were better answered in self-administered mode. However, overall test scores showed no significant mode differences. In another paper based on the same experiment, Gooch and Vavreck (2016) also found lower item non-response in the self-administered mode across most question types. Notably, mode effects were more pronounced among respondents with lower cognitive ability, who were more likely to skip items in in-person interviews.

The findings from these studies hold potential for application in large-scale surveys measuring cognitive abilities. On the one hand, the Army STARRS cognitive battery was successfully administered in a group setting without the need for an interviewer or assessor, suggesting its feasibility for adaptation to online self-administration. On the other hand, the experimental results from Gooch (2015) provide a further evidence of the potential for mode effects in cognitive scores, even for simple test batteries.

Table 7. Cognitive measures collected in the Army STARRS study (2014)

Domain	Test	Description
Executive function, Attention/ Concentration	Penn Conditional Exclusion Test (PCET)	Participants are asked to determine which object does not belong to a particular group of other objects. The objects vary on three characteristics: size, shape, and the thickness of the lines composing them.
Attention/ Concentration	Penn Continuous Performance Test (PCPT)	Participants are shown a series of configurations of red seven-segment displays and asked to press a space bar when the stimulus is a number (first half) or letter (second half).
Executive function	Short Letter-N-Back (SLNB)	Participants are asked to pay attention to letters that flash on the computer screen one at a time, and to press the spacebar whenever the letter on the screen is the same as the one before the previous letter (2-back).
Attention/ Concentration	Go/No-Go (GNG)	Participants see a series of Xs and Ys quickly displayed at different positions of the screen. They are instructed to respond (press the spacebar) if and only if an X appears in the upper half of the screen.
Working memory	Penn Face Memory Test (PFMT)	The test presents respondents 20 faces that they will be asked to identify later. After an initial learning period, they are shown a series of 40 faces (20 targets and 20 distractors) and are asked to decide whether they have seen each face before.

5.2 Innovations for the assessment of cognition

Self-administered cognitive assessments offer promising benefits for medical practice, particularly in enabling early diagnosis while reducing demands on clinical staff (Tsoy *et al.*, 2021). Advances in digital technology have led to the development of numerous self-administered tools, offering features such as automated scoring, efficient testing, increased sensitivity, and the potential to assess cognition in broader, more representative populations (Sternin *et al.*, 2019) These test batteries are assessed in various papers in the literature, including the reviews by Wild *et al.* (2008), Zygouris and Tsolaki (2014), Sternin *et al.* (2019) and Tsoy *et al.* (2021).

Significantly, the widespread use of smartphones (including among the ageing population) with increasing storage and connectivity abilities, has enabled the possibility of collecting large amounts of data with minor effort required from participants (Vasilichi, 2021). These apps can be advantageous in increasing engagement, reducing interviewer effects, increasing sample size, improving representativeness, and making participation more accessible and inclusive (Nicosia *et al.*, 2023). 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 (Vasilichi, 2021). Repeated measures can be helpful in identifying fluctuations in performance, which could be a meaningful metric of cognitive function, and a beneficial way to understand how context shapes cognitive performance (Weizenbaum *et al.*, 2020). There are countless examples of cognitive assessments collected via mobile apps and smartphones. We illustrate the data streams available for this purpose using the classification scheme suggested by Koo and Vizer (2019), which considers the following groups:

- **Game performance indicators**. Cognition has been assessed both with previously existing games (e.g. Thompson *et al.*, 2012), and with new especially created games designed to increase engagement while testing performance or delivering important content (e.g. Tong *et al.*, 2016; Pedersen *et al.*, 2023). Significant correlations between performance indicators and scores in conventional cognitive function assessments have been found for both types of game. The implementation of the Sea Hero Quest game app in the Understanding Society Innovation panel, described in Section 4.2, is a pioneering application in this field for probability-based surveys.

- **GPS data.** Trackers in smartphones and wearable sensors have been employed to identify the geographic area a person covers in daily life. In previous experiments (e.g. Tung *et al.*, 2014), distance covered has been identified as significantly different for elderly people with dementia or Alzheimer's disease, compared with the healthy control groups. A review of similar studies can be found in Cullen *et al.* (2022)

- **Activity daily-life performance**. Smartphone- and tablet-based apps have been designed to simulate daily life activities performed in a *virtual reality* setting including, for example, a virtual supermarket (e.g. Zygouris *et al.*, 2017; Tsai *et al.*, 2021), and simple tasks of facial recognition, face/name pairings, pillbox management, using an automated teller machine, and implementing an automated medical prescription using a telephone. As highlighted in the recent review by Veneziani *et al.* (2024), findings from these studies have demonstrated significant correlations between the scores obtained from traditional assessments and individual performance in these activities in virtual reality. This group can also include *wearable* technologies such as smartwatches, accelerometers, cameras, and glasses, both for cognitive monitoring and assistance. These technologies record time, location, temperature, and activity levels to create personalised profiles of risk, thus modelling behaviour and alerting caregivers when potentially dangerous events occur. A more detailed review of wearable technologies for cognitive assessment can be found in Vasilichi (2021).

- **Speech analysis apps**. Some vocal characteristics in speech have been used for cognitive assessment via smartphone apps, as various types of dementia and mild cognitive impairment can be manifested as irregularities in human speech and language. The app in Konig *et al.* (2018) records participants while performing short vocal cognitive tasks during a regular consultation. The voice recordings were processed using automatic speech processing and machine learning techniques. The app showed that the fluency and free speech tasks are highly accurate for automatic differentiation between mild cognitive impairments and Alzheimer's disease. Similar results have been reported in more recent applications, as reviewed in Al-Hammadi *et al.* (2024).

- **Physical movement analysis.** The link between fine motor skills in hand movements and cognitive impairments has been frequently reported in the literature (Ilardi *et al.*, 2022). Some apps have been designed to measure fine motor skills from tapping on a tablet screen, with analyses showing significant differences in finger dexterity between people with dementia and the healthy control group (Suzumura *et al.*, 2018).

In addition to these data streams, human interaction with smart homes can provide an accurate assessment of cognitive abilities (Vasilichi, 2021). Smart homes are fitted with a diverse network of sensors and advances using intelligent techniques, and offers support and responsible administration for grasping various benefits for its inhabitants (Javed *et al.*, 2021). Although such systems have not been designed to directly monitor condition, they can passively measure the ability of residents in executing simple to complex daily living activities, with machine learning algorithms applied to the data to conceptualise patterns of behaviour to assess cognitive health. The studies reviewed in Vasilichi (2021) and the Javed *et al.* (2021) demonstrate that these assessments can accurately identify mild cognitive impairments in participants.

To our knowledge, applications of these approaches in large scale surveys has been limited so far. One major challenge to their deployment is data privacy. Smartphone technology has enabled remote monitoring of health parameters such as physical activity and blood pressure, and these technologies are increasingly becoming familiar among the general public. However, cognitive assessments via phone apps are only feasible when participants provide explicit consent to data collection. Concerns about this information being shared, disclosed, or misused could potentially reduce the willingness to participate. Technological limitations also pose a significant barrier, as these systems depend on stable and fast internet connections, which can be difficult to maintain in certain settings.

Nevertheless, existing literature provides compelling evidence of the accuracy these methods can achieve in measuring cognitive abilities and enabling early diagnosis of cognitive impairments. Although not yet as reliable as traditional methods, mobile assessments demonstrate high levels of feasibility and validity, making them a promising tool for capturing individual cognitive variability in real-world contexts. An additional advantage of passive data collection is its ability to provide significantly higher temporal resolution and repeatability, both of which can greatly enhance the accuracy of cognitive assessments. Integrating passive and active cognitive data collection into large-scale, representative surveys represents a promising avenue for future research.

5.3 Survey response behaviour and cognitive measures

Recently, alternative approaches to infer cognitive abilities from other types of behaviour have been proposed. These approaches do not require the use of cognitive ability tests, and can overcome some of their practical limitations (Junghaenel *et al.*, 2023). Specifically, responding to a survey is an inherently complex and cognitively demanding task that requires attention, working memory, executive functioning, and short-and long-term memory (Jin *et al.*, 2023). Research suggests that examining survey response behaviour in the elderly population may represent a valuable resource that can be used to develop behaviour-based markers of cognitive decline that are cost-effective, unobtrusive, and scalable.

Jin *et al.* (2023) use two types of indices that summarise survey response behaviour to develop early markers of cognitive decline and dementia. Subtle reporting mistakes are derived from questionnaire answer patterns in several population-based longitudinal aging studies including HRS, ELSA, and the SHARE survey. More interestingly, they also analyse indices generated from computer use behaviours recorded on the backend server of the Understanding America web-based panel via paradata. The indices are grouped in four categories, all of which could be, according to the specialised literature, associated with cognitive impairment, namely:

- *Survey completion*, including aspects of survey completion, such as whether a survey is incomplete, and the time spent on completing a survey
- *Response time*, including median response times, variability, and intra-respondent correlation
- *Errors and corrective behaviours*, such as proportion of corrected or changed answers, rate of error messages received, and the rates of use of the “back” and “next” buttons.
- *Mouse and touch efficiency*, including median and variability of mouse clicks across screens.
- *Keystrokes*, measuring temporal rhythms of keystrokes from keyboard entries by respondents.

Preliminary findings, derived from a subset of indices, suggest that web-based survey paradata may hold promise for predicting cognitive decline and dementia. However, research is still ongoing, and conclusive results are not yet available. Junghaenel *et al.* (2023), analysing response times from over 6,000 respondents administered over 6.5 years in the Understanding America panel, find that the association between response time measures and cognitive assessments is relatively weak. Nevertheless, the study reveals that response time indicators exhibit a stronger association with cognitive assessments over lag periods ranging from one to at least six years, highlighting their potential utility for the prospective prediction of cognitive abilities.

Gao *et al.* (2024) analysed questionnaire response data from participants aged 50 years and older in waves 8 and 9 of the HRS (2006 and 2008). The authors generated low-quality response indices based on participant behaviour across four brief questionnaires, including factors such as skipped questions, contradictory answers, over-simplified responses, and inaccurate or unreliable responses. Using machine learning, they predicted cognitive status scores (measured in the same survey), as well as the incidence of dementia or mortality in the next ten years, derived from the HRS follow-up records. Their best-performing algorithm outperformed the efficiency of age or health-based screening strategies for identifying individuals at high risk of cognitive impairment.

Schneider *et al.* (2024) analysed data from ten epidemiological studies of ageing, including surveys such as ELSA, HRS, SHARE, and TILDA, all of which incorporate cognitive assessments. They derived six statistical indicators of survey response quality: item non-response, random measurement error, Guttman errors[8], multivariate outliers, acquiescent responses, and extreme responses. Their analysis showed a significant association between lower cognitive

[8] A Guttman error occurs when a respondent answers a more difficult question correctly but fails to correctly answer an easier one on the same scale, violating the expected pattern of a reliable survey (Guttman, 1944).

ability—particularly in processing speed and executive functioning—and reduced response quality.

These studies suggest that studying survey response behaviour could provide valuable information about cognitive functioning, especially for the elderly population. However, the association between indicators derived from survey response behaviour and formal cognitive assessments has been found to be relatively low. As a result, there is no indication that indicators derived from survey para-data could completely replace conventional cognitive assessments in terms of reliability or diagnostic accuracy. Instead, it may serve as a supplementary tool to enhance understanding or identify potential areas for further evaluation.

6 Conclusions, discussion, and recommendations

6.1 Conclusions

Online self-administration is increasingly being explored as a method for cognitive ability testing. Its appeal lies not only in the potential for significantly reducing fieldwork costs but also in its capacity to remove interviewer-led biases common in in-person surveys.

This evidence review has identified several large-scale surveys, predominantly longitudinal, that have successfully implemented self-administered online cognitive tests. While older adults may exhibit greater reluctance to online modes due to unfamiliarity with them, evidence from surveys such as the Health and Retirement Study (HRS) and the 1946 National Birth Cohort demonstrates successful adaptation of online cognitive testing for ageing populations. This approach is similarly viable for younger populations, with surveys making extensive use of validated online batteries like Cognitron, CogState, and TestMyBrain. Furthermore, major general population surveys such as Understanding Society confirm the overall feasibility of online cognitive assessment, irrespective of age.

However, the review also identifies significant mode effects when cognitive assessments are administered using different modes (e.g., in-person versus online self-administration). Such effects can compromise the comparability of scores both over time when a participant's mode of completion changes between waves, or across participants assessed in different modes at the same time point. These differences are particularly pronounced when tests originally designed for interviewer administration are adapted for online use and tend to be greater among respondents with lower cognitive functioning. To minimise these problems, it is therefore recommended to use test batteries specifically developed for online self-administration in mixed-mode surveys. In such cases, participants completing the study in person should also undertake the assessment via self-completion, ensuring that all respondents complete the test using the same mode.

Finally, the review examined innovative methods, such as using survey response behaviour as a proxy for cognitive function or employing online games and mobile applications for tracking cognitive abilities. While these approaches offer an attractive alternative to specialised tests, research in this area is still scarce. A significant challenge for these methods is that individual performance can be influenced by several confounding factors, including environmental conditions, the participant's physical state, and technical issues associated with devices and internet connections. Disentangling cognitive ability from these effects is essential to establish the reliability and validity of these innovative measures compared to conventional tests.

6.2 Discussion and recommendations for survey practice

Web-based assessments offer a scalable, cost-effective way to measure cognitive abilities in large populations, and their use in social research has expanded substantially. Cognition can be assessed through traditional tests adapted for self-administered online formats, as in the longitudinal and cohort studies reviewed, or through computerized batteries like Cognitron or CogState, and app-based smartphone assessments. Additional indicators, such as item non-response and response times, can offer further insights into cognitive function, particularly among older participants.

Although different cognitive assessments can yield varying results, scores for the same individual across tasks are usually highly correlated. This is attributed to the general cognitive ability factor (“g”), proposed by Spearman (1923) and consistently observed in factor analyses, where a strong first factor explains much of the variance in test scores. “g” is usually a strong predictor of real-world outcomes such as academic achievement, job performance, and health behaviours.

From a research design perspective, “g” suggests that general trends in cognitive ability can be captured by many test batteries. However, specific domains such as memory, attention, executive function and verbal reasoning contribute unique variance that is not captured by this factor alone, making careful battery selection essential.

Considering the issues addressed in this evidence review, we provide some recommendations regarding selecting web-based test battery in Section 6.2.1 and for tackling mode effects in the context of mixed mode surveys in Section 6.2.2.6.2.10

6.2.1 Selecting a test battery

In unimodal surveys, the issue of which test battery to select for cognitive function assessment is dependent on the objectives of the research project, and the domains for which the assessment is required. From a survey methodology perspective, taking advantage of the capabilities of each mode requires acknowledging their differences (Ofstedal *et al.*, 2021). For example, self-administered tests allow using visual stimuli, which is not possible in telephone interviews, while in-person tests including verbal communication, such as reading or repeating words, counting backwards, or naming animals, may be difficult to replicate in a web survey. On the other hand, web-based tests allow measuring domains such as processing speed or visuospatial orientation, which can be difficult to measure in conventional interviewer-led settings.

Age group is a key consideration. Cognitive ability, attention span, and technological familiarity vary across life stages. Tests for children are generally shorter and less complex; those for older adults often accommodate declines in processing speed, memory, and sensory abilities. In all cases, instructions should be simple and accessible - e.g. large fonts, audio support, clear language, intuitive icons.

Scientific validity and reliability of the test batteries might present significant issues. Tests should be validated independently, ensuring that they are reliable (internally consistent) and valid (accurate in measuring the intended constructs). They should also be sensitive enough to detect subtle differences in the cognitive domain being measured. Their suitability needs to be assessed case-by-case according to the objectives of each study. Cognitive assessments conducted by other modes and devices, such as those analysed in Section 5.2, have mostly been used in exploratory research. As they have not been deployed in large scale surveys, their psychometric properties and correlations with other cognitive assessments have not been studied in detail in the literature.

Practical feasibility of test administration is another issue to consider when selecting and implementing a web-based cognitive assessment battery. Easiness of administration is key to reduce withdrawals and incomplete responses. As demonstrated by the application of the Cognitron battery in the sample of 1946 National Birth Cohort participants (Cai *et al.*, 2024), completion rates can be high provided that intuitive and user-friendly interfaces with clear and simple instructions are provided, even for less technically literate population groups such as the elderly. Qualitative research conducted in this study yielded several practical recommendations for implementing online test batteries. Some of these include:

- Allowing participants to pause between tests or complete the assessments over multiple days, with an easy way of re-accessing the testing platform, to reduce potential stress.
- Facilitating platform 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.
- Informing participants that a stable internet connection is required and providing a phone or email helpline to address potential technical issues.
- Designing the test interface with clear language, intuitive icons, and a simplified consent form to accommodate respondents with lower technical literacy.
- Using larger font sizes, limiting flashing lights, and offering an audio option for written instructions to better facilitate participation by elderly users.

User testing can be a key determinant in the definition of layout and graphic design parameters, as the test batteries need to be accessible for all participants in the target population, and potentially across several devices (e.g. laptop, smartphone, tablet) (Wilson and Dickinson, 2022). For tests measuring processing speeds or reaction times, internet connection speed and reliability can also be a significant issue, especially in remote or low resource settings. Special measures to ensure connection stability might be required in these cases.

The costs of licensing batteries like Cognitron, Cogstate, or UK Biobank, or developing alternative methods to collect cognitive data, can be significant. Costs associated with any required software or computational platforms, technical support, and server hosting, should be accounted for when evaluating the feasibility of implementing these batteries in large-scale surveys. A reduced number of simple tests requiring verbal or text responses, might be helpful in collecting relevant information for several cognitive domains, while offering stronger chances of scalability at more manageable costs.

Finally, ensuring data confidentiality is essential to protect the privacy of participants and comply with legal and ethical standards. Participants require clear information about how the data collected will be used, stored, and protected. They also need to understand their rights, including the right to withdraw their data at any point, and the fact that their responses will be stored securely and not shared with third parties.

6.2.2 Tackling mode effects in cognitive testing

The issue of which test battery to implement and how to analyse the results is further complicated for mixed mode surveys due to the significant potential for mode effects. The results from experiments conducted within the HRS and Understanding Society demonstrate significant differences in cognitive assessment outcomes between participants completing the assessments in web-based self-administration and those completing them with the assistance (or presence) of an interviewer, with web respondents generally performing better. In both cases, the tests used were adaptations of “gold standard” in-person measures adapted for online administration. Factors such as a controlled and familiar environment, reduced social desirability bias, enhanced focus, and fewer distractions and cognitive load, have been proposed as potential explanations for these differences. Although unlikely to account for much of the differences, some level of cheating amongst web respondents cannot be ruled out. The relationship between participant engagement and task difficulty might also be another confounding factor.

Selection biases may play a role, as participants who choose to respond to surveys via web-based modes often have higher levels of education and digital literacy compared to the general population. This selection bias can contribute to the better performance in cognitive assessments among web respondents. However, the findings in Al Baghal (2019) suggest that measurement mode effects play a more substantial role in explaining outcome differences than mode self-selection due to cognitive abilities. This poses a significant challenge as measurement differences cause difficulties in interpreting differences both across and within respondents. Regardless of the motivating factors, our evidence review finds consistent evidence of mode effects across surveys and cognitive batteries, especially for test batteries that were originally designed for in-person administration. The extent to which these mode differences are relevant depends on the specific research question of the study. However, as reported by Smith *et al.* (2023), “*even subtle effects attributable to mode can be comparable in magnitude to effects of risk factors important to population health*” (p. 197), which highlights the need to identify them and account for them in substantive analyses. Importantly, mode differences are likely to be greatest for respondents with lower cognitive functioning levels (Ofstedal *et al.*, 2021; Smith *et al.*, 2023), which can overwhelm substantive findings on studies focusing on this dimension.

As comparability between modes is a priority for mixed-mode surveys, it is important to identify tests that are suitable for administration across modes, and to make the necessary adjustments to reduce the chance of mode effects. Some of these adjustments could include:

- Prioritising measures designed for self-completion (especially those developed for online use) and adapt them for other modes as needed. This is preferable to adapting interviewer-led “gold standard” measures for self-completion.
- Choosing simpler tasks that require short responses, as they are potentially less sensitive to mode effects, compared to more complex tasks that are highly dependent on specific modes or devices (in the case of web administration).
- Ensuring that instructions for cognitive tasks are consistent across modes, that they use scripts and simplified language to minimise variability, and that the same time limits are imposed for each task regardless of mode.
- Encouraging participants to complete assessments in quiet environments that reduce the possibility of distractions.

Reminding participants that honest answers are required, and no one is expected to get a perfect score could be a good recommendation if cheating is a concern (Lachman and Alwin, 2008).However, regardless of how comparable the tests are in terms of administration protocol, measurement differences by mode are still likely to be present. This means that careful attention to mode differences will be required when analysing the resulting data (Ofstedal *et al.*, 2021). Some possible courses of action to deal with cognitive data from different modes in models of cognitive function are discussed in Smith *et al.* (2023), and include: using mode as a covariate, standardising scores to ensure consistency of means and standard deviations across modes, and collapsing continuous scores into categories to reduce measurement biases by mode. Still, these measures are unlikely to solve the possible issues of confounding effects when comparing measures obtained using different modes.

When web modes are introduced in a mixed-mode longitudinal survey, variations in cognitive scores, both within and across participants, are likely to occur. These variations are often at least partially attributable to the change in the mode of administration, which can significantly disrupt the continuity of cognitive trajectories over time. The shift to mixed mode may also necessitate introducing new assessments that are better suited to online administration. This presents a key challenge and potential barrier, in particular for ageing studies considering a move from in-person interviewing to a web-first mixed mode approach. Calibration studies, in which participants complete both the legacy measures and the newly introduced ones, may be needed for quantifying mode effects and testing adjustment strategies. Although demanding in terms of cost and logistics, such studies may be beneficial to safeguard comparability across waves and help enable transitions to web-first data collection.

6.3 Recommendations for further research

Cognitive testing conducted in person has a well-established history in survey research, supported by extensive evidence on its reliability, validity, and long-term trends in cognitive performance across populations. In comparison, online self-administered tests are a more recent development, and the evidence base on their use is consequently more limited. This evidence review strongly suggests that cognitive tests designed for online administration should be preferred over tests adapted from interviewer-administered formats, as findings point to significant mode effects in the latter. In the case of mixed-mode surveys, those participating in-person should complete assessments via self-completion, although even the presence of an interviewer can still have an impact on performance. Further research is needed to identify the precise causes of score differences across modes and to develop effective mitigation strategies that ensure consistent response behaviour. Experimental data can be crucial for studying these aspects. A related and critical question for longitudinal research, where consistent time-series data are essential, is how to statistically adjust for mode effects when analysing cognitive data collected in different modes over time

Similarly, the effects of the specific device used for online self-administered tests have yet to be studied in detail. Although limited, the existing evidence suggests that device characteristics such as latency (the delay between when a user takes an action and when they get a response from the device), screen size, input method and operating system can influence cognitive scores. These device factors are often confounded with sociodemographic characteristics, further complicating cross-device score comparisons. Selecting appropriate tasks for each cognitive dimension and statistically accounting for the device used are two measures to reduce this bias. However, further research is required to establish best practices.

Finally, non-conventional approaches (such as online games, indicators derived from survey response behaviour, and data from mobile apps) show great potential for measuring cognitive function in online surveys. However, their reliability and validity have not been properly assessed. A more fundamental challenge is that these measures are not dimension-specific, and their results likely incorporate factors related to individual and contextual characteristics, which may be confounded with cognitive ability. Developing methods to disentangle these effects and resolve this discrepancy is a crucial direction for future research.

References

Al-Hammadi, M., Fleyeh, H., Åberg, A.C., Halvorsen, K. and Thomas, I. (2024) 'Machine learning approaches for dementia detection through speech and gait analysis: A systematic literature review', *Journal of Alzheimer's Disease*, 100(1), p. 27.

Al Baghal, T. (2019) 'The effect of online and mixed-mode measurement of cognitive ability', *Social Science Computer Review*, 37(1), pp. 89-103.

Allen, K.L., Byrne, S.M., Hii, H., van Eekelen, A., Mattes, E. and Foster, J.K. (2012) 'Neurocognitive functioning in adolescents with eating disorders: A population-based study', *Cognitive Neuropsychiatry*, 18(5), pp. 355-375.

Alzheimer's Research UK (2025) *Mild Cognitive Impairment*. Available at: https://www.alzheimersresearchuk.org/dementia-information/types-of-dementia/mild-cognitive-impairment/ (Accessed: 20 October 2025).

Bleicher, K., Summerhayes, R., Baynes, S., Swarbrick, M., Cristina, T.N., Luc, H., Dawson, G., Cowle, A., Dolja-Gore, X. and McNamara, M. (2023) 'Cohort Profile Update: The 45 and Up Study', *International Journal of Epidemiology*, 52(1), pp. e92-e101.

Brody, D.J., Kramarow, E.A., Taylor, C.A. and McGuire, L.C. (2019) 'Cognitive performance in adults aged 60 and over: National Health and Nutrition Examination Survey, 2011–2014', *National Health Statistics Reports*, 126, pp. 1-23.

Burton, J., Jäckle, A. and Couper, M. (2024) 'The effect of incentives on the willingness to participate in a gamified app study', in Vine, J.e.a. (ed.) *Understanding Society Working Paper Series 2024-11*. Colchester, Essex: Institute for Social and Economic Research, University of Essex.

Cai, Z., Giunchiglia, V., Street, R., Del Giovane, M., Lu, K., Popham, M., Wong, A., Murray-Smith, H., Richards, M., Crutch, S., Hellyer, P.J., Schott, J.M. and Hampshire, A. (2024) 'Online46: online cognitive assessments in elderly cohorts - the British 1946 birth cohort case study', *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 17(2), p. e70098.

Chan, T., Fruiht, V. and McInnis, N. (2024) 'Building blocks for a happy life: Longitudinal associations between early life income, mentorship and later well-being', *American Journal of Community Psychology*, 74(1-2), pp. 16-30.

Coutrot, A., Al Baghal, T. and Spiers, H. (2024) 'Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16', in Vine, J.e.a. (ed.) *Understanding Society Working Paper Series 2024-11*. Colchester, Essex: Institute for Social and Economic Research, University of Essex.

Coutrot, A., Ricardo, S., Manley, E., de Cothi, W., Sami, S., Bohbot, V.D., Wiener, J.M., Hölscher, C., Dalton, R.C., Hornberger, M. and Spiers, H.J. (2018) 'Global determinants of navigation ability', *Current Biology*, 28(17), pp. 2861-2866.

Cullen, A., Mazhar, M.K.A., Smith, M.D., Lithander, F.E., Breasail, M.Ó. and Henderson, E.J. (2022) 'Wearable and portable GPS solutions for monitoring mobility in dementia: A systematic review', *Sensors*, 22(9), p. 3336.

Cummins, H., Bransby, L., Fredrickson, A., Stellman, G., Maruff, P., Edgar, C.J. and Lim, Y.Y. (2025) 'Acceptability and validity of the smartphone-administered version of the Cogstate Brief Battery in middle-aged and older adults of the Healthy Brain Project', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 20(S3), p. e091946.

de Leeuw, E.D. (2005) 'To mix or not to mix data collection modes in surveys', *Journal of Official Statistics*, 21(2), pp. 233-255.

Del Giovane, M., Giunchiglia, V., Cai, Z., Leoni, M., Street, R., Lu, K., Wong, A., Popham, M., Nicholas, J.M., Trender, W., Helleyer, P.J., Parker, T.D., Murray-Smith, H., Malhotra, P.A., Crutch, S.J., Richards, M., Hampshire, A. and Schott, J.M. (2025) 'Remote cognitive tests predict neurodegenerative biomarkers in the Insight 46 cohort', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 21(2), p. e14572.

Diamond, A. (2013) 'Executive functions', *Annual Review of Psychology*, 64, pp. 135-168.

Domingue, B.W., McCammon, R.J., West, B.T., Langa, K.M., Weir, D.R. and Faul, J. (2023) 'The mode effect of Web-based surveying on the 2018 U.S. health and retirement study measure of cognitive functioning', *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 78(9), pp. 1466-1473.

Edgar, C.J. (2023) 'Adaptation and modification of digital cognitive assessments for smartphone-based and unsupervised conduct', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 19(S21), p. e078218.

Fawns-Ritchie, C. and Deary, I.J. (2020) 'Reliability and validity of the UK Biobank cognitive tests', *PLos ONE*, 15(4), p. e0231627.

Folstein, M.F., Folstein, S.E. and McHugh, P.R. (1975) '“Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician', *Journal of Psychiatric Research*, 12(3), pp. 189-198.

Freedman, V.A. (2017) *The Panel Study of Income Dynamics' Well Being and Daily Life Supplement (PSID-WB) User Guide: Final Release 1*. Institute for Social Research, University of Michigan. Available at: https://psidonline.isr.umich.edu/WB/WBUserGuide.pdf (Accessed: 20 January 2025).

G2Aging (2024) *Health & Retirement Studies - Overview*. Gateway to Global Aging Data. Available at: https://g2aging.org/survey/overview (Accessed: 5 November 2024).

Gao, H., Schneider, S., Hernandez, R., Harris, J., Maupin, D., Junghaenel, D., Kapteyn, A., Arthur, S., Zelinski, E., Meijer, E., Lee, P.-J., Orriens, B. and Jin, H. (2024) 'Early identification of cognitive impairment in community environments through modeling subtle inconsistencies in questionnaire responses: Machine learning model development and validation', *JMIR Formative Research*, 8, p. e54335.

Giunchiglia, V., Cai, Z., Del Giovane, M., Street, R.E., Lu, K., Wong, A., Popham, M., Murray-Smith, H., Richards, M., Crutch, S.J., Schott, J.M. and Hampshire, A. (2025) 'Usability and compliance of online cognitive assessments in the British 1946 Birth Cohort', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 20(S3), p. e091632.

Gnambs, T. (2019) *NEPS Technical Report for English Reading Competence: Scaling Results of Starting Cohort 5 (Wave 12)* (NEPS Survey Paper No. 102). Leibniz Institute for Educational Trajectories, National Educational Panel Study. Available at: https://doi.org/10.5157/NEPS:SP53:1.0 (Accessed: 4 November 2024).

Golding, J., Pembrey, M., Jones, R. and Team, A.S. (2001) 'ALSPAC - The Avon Longitudinal Study of Parents and Children I. Study Methodology', *Pediatric and Perinatal Epidemiology*, 15, pp. 74-87.

Gomensoro, A. and Meyer, T. (2017) 'TREE (Transitions from Education to Employment): A Swiss multi-cohort survey', *Longitudinal and Life Course Studies*, 8(2), pp. 209-224.

Gooch, A. (2015) 'Measurements of cognitive skill by survey mode: Marginal differences and scaling similarities', *Research and Politics*, 2(3).

Gooch, A. and Vavreck, L. (2016) 'How face-to-face interviews and cognitive skill affect item non-response: A randomized experiment assigning mode of interview', *Political Science Research and Methods*, 7(1), pp. 143-162.

Guttman, L. (1944) 'A basis for scaling qualitative data', *American Sociological Review*, 9(2), pp. 139-150.

Hampshire, A., Azor, A., Atchison, C., Trender, W., Hellyer, P.J., Giunchiglia, V., Husain, M., Cooke, G.S., Cooper, E., Lound, A., Donnelly, C.A., Chadeau-Hyam, M., Ward, H. and Elliott, P. (2024) 'Cognition and memory after Covid-19 in a large community sample', *The New England Journal of Medicine*, 390(9), pp. 806-818.

Heller, K.A. and Perleth, C. (2000) *Cognitive Ability Test for Grades 4 to 12, Revision*. 1st Edition edn. Göttingen: Beltz.

Hsieh, S., McGrory, S., Leslie, F., Dawson, K., Ahmed, S., Butler, C.R., Rowe, J.B., Mioshi, E. and Hodges, J.R. (2014) 'The Mini-Addenbrooke's Cognitive Examination: A new assessment tool for dementia ', *Dementia and Geriatric Cognitive Disorders*, 39(1-2), pp. 1-11.

Ilardi, C.R., Iavarone, A., La Marra, M., Iachini, T. and Chieffi, S. (2022) 'Hand movements in Mild Cognitive Impairment: clinical implications and insights for future research', *Journal of Integrative Neuroscience*, 21(2), p. 67.

Javed, A.R., Fahad, L.G., Farhan, A.A., Abbas, S., Srivastava, G., Parizi, R.M. and Khan, M.S. (2021) 'Automated cognitive health assessment in smart homes using machine learning', *Sustainable Cities and Society*, 65, p. 102572.

Jin, H., Junghaenel, D.U., Orriens, B., Lee, P.-J. and Schneider, S. (2023) 'Developing early markers of cognitive decline and dementia derived from survey response behaviors: Protocol for analyses of preexisting large-scale longitudinal data', *JMIR Research Protocols*, 12, p. e44627.

Johnson, D., McGonagle, K., Freedman, V. and Sastry, N. (2019) 'Fifty years of the Panel Study of Income Dynamics: Past, present, and future', *The ANNALS of the American Academy of Political and Social Science*, 680(1), pp. 9-28.

Junghaenel, D.U., Schneider, S., Orriens, B., Jin, H., Lee, P.-J., Kapteyn, A., Meijer, E., Zelinski, E., Hernandez, R. and Stone, A.A. (2023) 'Inferring cognitive abilities from response times to Web-administered survey items in a population-representative sample', *Journal of Intelligence*, 11(3), pp. 1-25.

Kantar Public and National Centre for Social Research (NatCen) (2022) *Understanding Society: Wave 14 technical report*. The Institute for Social and Economic Research. Available at: https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/main-survey/technical-reports/6614-main-survey-technical-report-w14.pdf (Accessed: 1 August 2024).

Kapteyn, A., Angrisani, M., Darling, J. and Gutsche, T. (2024) 'The Understanding America Study (UAS)', *BMJ Open*, 14, p. e088183.

Kapteyn, A., Darling, J., Gatz, M., Orriens, B., Meijer, E., Liu, Y., Gutsche, T. and Schneider, S. (2021) 'Developing web-based cognitive functioning measures in the Understanding America Study'. Workshop: Measuring Cognition in Online Surveys, Generations and Gender Programme. Available at: https://www.ncrm.ac.uk/research/genpopweb2/events.php (Accessed: 20 January 2025).

Knyspel, J. and Plomin, R. (2024) 'Comparing factor and network models of cognitive abilities using twin data', *Intelligence*, 104, p. 101833.

Kobayashi, L.C. and Feldman, J.M. (2019) 'Employment trajectories in midlife and cognitive performance in later life: longitudinal study of older American men and women', *Journal of Epidemiology & Community Health*, 79, pp. 232-238.

Konig, A., Satt, A., Sorin, A., Hoory, R., Derreumaux, A., David, R. and Robert, P.H. (2018) 'Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people', *Current Alzheimer Research*, 15(2), pp. 120-129.

Koo, B.M. and Vizer, L.M. (2019) 'Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review ', *Innovation in Aging*, 3(1), p. igy038.

Kooijman, M.N., Kruithof, C.J., van Duijn, C.M., Duijts, L., Franco, O.H., van IJzendoorn, M.H., de Jongste, J.C., Klaver, C.C.W., van der Lugt, A., Mackenbach, J.P., A. Moll, H., Peeters, R.P., Raat, H., Rings, E.H.H.M., Rivadeneira, F., van der Schroeff, M.P., Steegers, E.A.P., Tiemeier, H., Uitterlinden, A.G., Verhulst, F.C., Wolvius, E., Felix, J.F. and Jaddoe, V.W.V. (2017) 'The Generation R Study: design and cohort update 2017', *European Journal of Epidemiology*, 31, pp. 1243-1264.

Krebs-Oesch, D., Jann, B. and Hupka-Brunner, S. (2023) *Implementation of a cognitive ability test in the TREE2 panel survey* (TREE Technical Paper Series No. 4). TREE (Transitions from Education to Employment), University of Bern, Switzerland. Available at: https://doi.org/10.48350/183109 (Accessed: 4 November 2024).

Lachman, M.E., Agrigoroaei, S., Tun, P.A. and Weaver, S.L. (2013) 'Monitoring cognitive functioning: Psychometric properties of the Brief Test of Adult Cognition by Telephone (BTACT)', *Assessment*, 21(4), pp. 404-417.

Lachman, M.E. and Alwin, D. (2008) 'Cognitive testing in large-scale surveys: assessment by telephone', in Hofer, S.M. and Alwin, D.F. (eds.) *Cognitive testing in large-scale surveys: Assessment by telephone*. SAGE Publications, Inc.

Lennartsson, C., Agahi, N., Hols-Salén, L., Kelfve, S., Kåreholt, I., Lundberg, O., Parker, M.G. and Thorslun, M. (2014) 'Data Resource Profile: The Swedish Panel Study of Living Conditions of the Oldest Old (SWEOLD)', *International Journal of Epidemiology*, 433(3), pp. 731-738.

Liu, Y., Schneider, S., Orriens, B., Meijer, E., Darling, J., Gutsche, T. and Gatz, M. (2022) 'Self-administered Web-based tests of executive functioning and perceptual speed:Measurement development study with a large probability-based survey panel', *Journal of Medical Internet Research*, 24(5), p. e34347.

Lockhart, C., Bright, J., Ahmadzadeh, Y., Breen, G., Bristow, S., Boyd, A., Downs, J., Hotopf, M., Palaiologou, E., Rimfeld, K., Maxwell, J., Malanchini, M., McAdams, T.A., McMillan, A., Plomin, R. and Eley, T.C. (2023) 'Twins Early Development Study (TEDS): A genetically sensitive investigation of mental health outcomes in the mid-twenties', *JCPP Advances*, 3(2), p. e12154.

Lupu, T., Braw, Y., Sacher, Y. and Ratmansky, M. (2021) 'Cogstate Brief Battery: Cognition and the feigning of cognitive impairment in chronic pain', *Applied Neuropsychology: Adult*, 29(6), pp. 1332-1343.

Malanchini, M., Rimfeld, K., Gidziela, A., Cheesman, R., Allegrini, A.G., Shakeshaft, N., Schofield, K., Packer, A., Ogden, R., McMillan, A., Ritchie, S.J., Dale, P.S., Eley, T.C., von Stumm, S. and Plomin, R. (2021) 'Pathfinder: a gamified measure to integrate general cognitive ability into the biological, medical, and behavioural sciences', *Molecular Psychiatry*, 26, pp. 7823-7837.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P. and Pietrzak, R.H. (2009) 'Validity of the CogState Brief Battery: Relationship to Standardized Tests and Sensitivity to Cognitive Impairment in Mild Traumatic Brain Injury, Schizophrenia, and AIDS Dementia Complex', *Archives of Clinical Neuropsychology*, 24(2), pp. 165-178.

McFall, S. (2013) *Understanding Society: UK Household Longitudinal Study: Cognitive ability measures*. Institute for Social and Economic Research, University of Essex. Available at: https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/user-guides/6614-user-guide-cognitive-ability-measures.pdf (Accessed: 12 December 2025).

Mielke, M.M., Machulda, M.M., Hagen, C.E., Edwards, K.K., Roberts, R.O., Pankratz, V.S., Knopman, D.S., Jack, C.R. and Petersen, R.C. (2015) 'Performance of the CogState computerized battery in the Mayo clinic study on aging', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 11(11), pp. 1367-1376.

Moore, T.M., Gur, R.C., Thomas, M.L., Brown, G.G., Nock, M.K., Savitt, A.P., Keilp, J.G., Heeringa, S., Ursano, R.J., Stein, M.B. and On behalf of the Army STARRS Collaborators (2019) 'Development, administration, and structural validity of a brief, computerized neurocognitive battery: Results from the Army Study to Assess Risk and Resilience in Servicemembers', *Assessment*, 26(1), pp. 3-143.

NHS England (2024) *Overview: Alzheimer's Disease*. National Health Service. Available at: https://www.nhs.uk/conditions/alzheimers-disease/ (Accessed: 20 October 2025).

Nicosia, J., Aschenbrenner, A.J., Balota, D.A., Sliwinski, M.J., Tahan, M., Adams, S., Stout, S.S., Wilks, H., Gordon, B.A., Benzinger, T.L.S., Fagan, A.M., Xiong, C., Bateman, R.J., Morris, J.C. and Hassenstab, J. (2023) 'Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease', *Journal of the International Neuropsychological Society*, 29(5), pp. 459-471.

Ofstedal, M.B., Fisher, G.G. and Herzog, A.R. (2005) *Documentation of Cognitive Functioning Measures in the Health and Retirement Study* (HRS Documentation Report DR-006). Survey Research Center, University of Michigan, Ann Arbor MI.

Ofstedal, M.B., McClain, C.A. and Couper, M.P. (2021) 'Measuring cognition in a multi-mode context', in Lynn, P. (ed.) *Advances in Logitudinal Survey Methodology*. John Wiley & Songs, Ltd., pp. 250-271.

Parker, R.M., Baker, D.W., Williams, M.V. and Nurss, J.R. (1995) 'The test of functional health literacy in adults', *Journal of General Internal Medicine*, 10, pp. 537-541.

Passell, E., Strong, R.W., Rutter, L.A., Kim, H., Scheuer, L., Martini, P., Grinspoon, L. and Germine, L. (2021) 'Cognitive test scores vary with choice of personal digital device', *Behavior Research Methods*, 53, pp. 2544-2557.

Patterson, J. (2011) 'F-A-S Test', in Kreutzer, J.S., DeLuca, J. and Caplan, B. (eds.) *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NS.

Pedersen, M.K., Castaño Díaz, C.M., Wang, Q.J., Alba-Marrugo, M.A., Amidi, A., Basaiawmoit, R.V., Bergenholtz, C., Christiansen, M.H., Gajdacz, M., Hertwig, R., Ishkhanyan, B., Klyver, K., Ladegaard, N., Mathiasen, K., Parsons, C., Rafner, J., Villadsen, A.R., Wallentin, M., Zana, B. and Sherson, J.F. (2023) 'Measuring cognitive abilities in the wild: Validating a population-scale game-based cognitive assessment', *Cognitive Science: A Multidisciplanry Journal*, 47(6), p. e13308.

Ramos-Serrano, R. and Fors, S. (2024) *SWEOLD 2014-2021: Analysis of trend shifts, non response, and attrition* (SWEOLD). Stockholm: Aging Research Centre. Available at: https://www.sweold.se/wp-content/uploads/2024/09/Serrano-Fors-2024-SWEOLD-NRA-REPORT.pdf (Accessed: 7 November 2024).

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., Dale, P.S. and Eley, T.C. (2019) 'Twins Early Development Study: A genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood', *Twin Research and Human Genetics*, 22(6), pp. 508-513.

Schneider, S., Lee, P.-J., Hernandez, R., Junghaenel, D.U., Stone, A.A., Meijer, E., Jin, H., Kapteyn, A., Orriens, B. and Zelinski, E.M. (2024) 'Cognitive functioning and the quality of survey responses: An individual participant data meta-analysis of 10 epidemiological studies of aging', *The Journals of Gereontology, Series B: Psychological Sciences and Social Sciences*, 79(5), p. gbae030.

Singh, S., Strong, R.W., Jung, L., Haofei Li, F., Grinspoon, L., Scheuer, L.S., Passell, E.J., Martini, P., Chaytor, N., Soble, J.R. and Germine, L. (2021) 'The TestMyBrain digital neuropsychology Toolkit: Development and psychometric characteristics', *Journal of Clinical and Experimental Neuropsychology*, 43(8), pp. 786-795.

Smith, J.R., Gibbons, L.E., Crane, P.K., Mungas, D.M., Glymour, M.M., Manly, J.J., Zahodne, L.B., Mayeda, E.R., Jones, R.N. and Gross, A.L. (2023) 'Shifting of cognitive assessments between face-to-face and telephone administration: Measurement considerations', *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 78(2), pp. 191-200.

Sonnega, A., Faul, J., Ofstedal, M.B., Langa, K.M., Phillips, J.W.R. and Weir, D.R. (2014) 'Cohort Profile: the Health and Retirement Study (HRS)', *International Journal of Epidemiology*, 43(2), pp. 576-585.

Spearman, C. (1923) *The Nature of "Intelligence" and the Principles of Cognition*. London.

Spiers, H.J., Coutrot, A. and Hornberger, M. (2023) 'Explaining world-wide variation in navigation ability from millions of people: Citizen science project Sea Hero Quest', *Topics in Cognitive Science*, 15(1), pp. 120-138.

Sternin, A., Burns, A. and Owen, A.M. (2019) 'Thirty-five years of computerized cognitive assessment of aging—Where are we now?', *Diagnostics*, 9, p. 114.

Stricker, N.H., Lundt, E.S., Edwards, K.K., Machulda, M.M., Kremers, W.K., Roberts, R.O., Knopman, D.S., Petersen, R.C. and Mielke, M.M. (2019) 'Comparison of PC and iPad administrations of the Cogstate Brief Battery in the Mayo Clinic Study of Aging: Assessing cross-modality equivalence of computerized neuropsychological tests', *The Clinical Neuropsychologist*, 33(6), pp. 1102-1126.

Suzumura, S., Osawa, A., Maeda, N., Sano, Y., Kandori, A., Mizuguchi, T., Yin, Y. and Kondo, I. (2018) 'Differences among patients with Alzheimer's disease, older adults with mild cognitive impairment and healthy older adults in finger dexterity', *Geriatrics and Gerontology*, 18(6), pp. 907-914.

Thompson, O., Barrett, S., Patterson, C. and Craig, D. (2012) 'Examining the neurocognitive validity of commercially available, smartphone-based puzzle games', *Psychology*, 3(7), pp. 525-526.

Tong, T., Chignell, M., Tierney, M.C. and Lee, J. (2016) 'A serious game for clinical assessment of cognitive status: Validation study', *JMIR Serious Games*, 27(4(1)), p. e7.

Trouton, A., Spinath, F.M. and Plomin, R. (2002) 'Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood', *Twin Research*, 5(3), pp. 448-448.

Tsai, C.-F., Chen, C.-C., Hsiao-Kuang Wu, E., Chung, C.-R., Huang, C.-Y. and Tsai, P.-Y. (2021) 'A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, pp. 2124-2132.

Tsigaridis, K., Gaia, A., Moulton, V., Wright, L., Brown, M. and Silverwood, R. (2025) 'Evaluating mode effects on a cognitive assessment administered via face-to-face, Web and video interviewing: Insights from new experimental data', *11th European Social Research Association (ESRA) Conference*. Utrecht, The Netherlands, July 14-18, 2025.

Tsoy, E., Zygouris, S. and Possin, K.L. (2021) 'Current state of self-administered brief computerized cognitive assessments for detection of cognitive disorders in older adults: A systematic review', *The Journal of Prevention of Alzheimer's Disease*, 8, pp. 267-276.

Tung, J.Y., Rose, R.V., Gammada, E., Lam, I., Roy, E.A., Black, S.E. and Poupart, P. (2014) 'Measuring life space in older adults with mild-to-moderate Alzheimer's Disease using mobile phone GPS', *Geriatrics and Gerontology*, 60(2), pp. 154-162.

Ursano, R.J., Colpe, L.J., Heeringa, S.G., Kessler, R.C., Schoenbaum, M., Stein, M.B. and On behalf of the Army STARRS collaborators (2014) 'The Army Study to Assess Risk and Resilience in Service members (Army STARRS)', *Psychiatry: Interpersonal and Biological Processes*, 77(2), pp. 107-119.

Vasilichi, A. (2021) *Measuring cognition in population-based cohort studies*. UCL Centre for Longitudinal Studies. Available at: https://www.ukri.org/wp-content/uploads/2019/07/ESRC-220311-MeasuringCognitionInPopulationBasedCohortStudies-190705.pdf (Accessed: 18 December 2024).

Veneziani, I., Marra, A., Formica, C., Grimaldi, A., Marino, S., Quartarone, A. and Maresca, G. (2024) 'Applications of artificial intelligence in the neuropsychological assessment of dementia: A systematic review', *Journal of Personalized Medicine*, 14(1), p. 113.

von Stumm, S., Kandaswamy, R. and Maxwell, J. (2023) 'Gene-environment interplay in early life cognitive development', *Intelligence*, 98, p. 101748.

Wadsworth, M., Kuh, D., Richards, M. and Hardy, R. (2006) 'Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development)', *International Journal of Epidemiology*, 35(1), pp. 49-54.

Wechsler, D. (1981) *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York, US: Psychological Corporation.

Weizenbaum, E., Torous, J. and Fulford, D. (2020) 'Cognition in context: Understanding the everyday predictors of cognitive performance in a new era of measurement', *JMIR mHealth and uHealth*, 8(7).

Wild, K., Howieson, D., Webbe, F., Seelye, A. and Kaye, J. (2008) 'The status of computerized cognitive testing in aging: A systematic review', *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 4(6), pp. 428-437.

Wilson, L. and Dickinson, E. (2022) *Respondent Centred Surveys: Stop, Listen and Then Design*. London: SAGE Publications Ltd.

Woodcock, R.W. and Johnson, M.B. (1989) *Woodcock-Johnson Tests of Cognitive Ability*. DLM Teaching Resources.

Wu, A.F.-W., Henderson, M., Brown, M., Adali, T., Silverwood, R.J., Peycheva, D. and Calderwood, L. (2024) 'Cohort Profile: Next Steps—the longitudinal study of people in England born in 1989–90 ', *International Journal of Epidemiology*, 53(6), p. dyae152.

Zhuang, L., Yang, Y. and Gao, J. (2021) 'Cognitive assessment tools for mild cognitive impairment screening', *Journal of Neurology*, 268, pp. 1615-1622.

Zinn, S., Landrock, U. and Gnambs, T. (2021) 'Web-based and mixed-mode cognitive large-scale assessments in higher education: An evaluation of selection bias, measurement bias, and prediction bias', *Behavior Research Methods*, 53, pp. 1202-1217.

Zygouris, S., Ntovas, K., Giakoumis, D., Votis, K., Doumpoulakis, S., Segkouli, S., Karagiannidis, C., Tzovaras, D. and Tsolaki, M. (2017) 'A preliminary study on the feasibility of using a virtual reality cognitive training application for remote detection of mild cognitive impairment', *Journal of Alzheimer's Disease*, 56(2), pp. 619-627.

Zygouris, S. and Tsolaki, M. (2014) 'Computerized cognitive testing for older adults: A review', *American Journal of Alzheimer's Disease & Other Dementias*, 30(1), pp. 13-28.

www.surveyfutures.net