A large public dataset of computerized cognitive test results with high-resolution speech quantification

Michael Blank, PhD[1], Kathleen Hall, PhD[1], Kristin Geraci, PhD[1], David K Johnson, PhD[2], Peter Pebler, M.S.[1], Isabella Jaramillo[1], Omar Kahly[1] and David L. Woods, PhD[1]
(1) Neurobehavioral Systems, Inc, Berkeley, CA, USA; (2) UC Davis Alzheimer's Disease Center, Walnut Creek, CA, USA

Introduction

We present a publicly available normative dataset with results from the California Cognitive Assessment Battery (CCAB) [1], to support research into cognition, aging, and speech and language biomarkers. This dataset is hosted on Open Science Framework (OSF) to facilitate collaborative and/or independent secondary analyses by other researchers.

Methods

Participants: 1,,867 participants completed the CCAB normative protocol (55.4% female; 36.3% white; mean age 53.5 years ± 17.3 years; see demographics table for more information).

Protocol: Most participants completed ~6 hours of supervised testing across 3 consecutive days (identical tests were administered on days 2 & 3); others (N=827) completed a single 2-hour test session. Some participants have completed follow-up assessments at one (N=587), two (N=253), and three (N=187) years post-enrollment. More than 95% of participants were tested in their homes with telemedical proctoring.

Technology: CCAB is an automated, computerized test battery with 22 verbal and 18 non-verbal tasks, administered on calibrated hardware with Presentation® software for sub-millisecond temporal precision. CCAB tasks span multiple cognitive domains and response modalities, including touch (e.g., trail making), mouse (e.g., choice reaction time), and speech (e.g., picture description). Extended demographic and health data were also collected, including psychological scales (e.g., GDS, CFQ, GAD-7), a vocabulary-based estimate of premorbid intelligence, and lifestyle and medical history. For a complete list of tests and descriptions, visit ccabresearch.com

Verbal responses were recorded, then transcribed using consensus automatic speech recognition (CASR), which combines multiple ASR engines to achieve >98% word-level transcription accuracy. High-quality recordings and high-accuracy automated transcriptions enable fine-grained quantification of speech timing, fluency, syntax, and phonetics.

Analysis: Each individual test run (N > 120,000) was analyzed to produce summary performance metrics such as category-wise response counts, speech complexity measures and acoustics, and response timing. Results were collated by task, deidentified, and paired with metadata and data dictionaries following OSF best practices.

Results

Age	Count	Years Education (SD)	Gender (% Female)	Non-hispanic white (%)	Latino (%)	Black (%)	Asian (%)
18-29	224	14.7 (2.5)	58.0	21.0	29.0	16.5	25.4
30-39	288	15.2 (2.8)	52.1	26.4	21.5	18.1	20.8
40-49	191	15.4 (2.8)	53.4	30.4	20.4	15.7	19.9
50-59	345	15.2 (2.9)	54.8	33.6	16.5	26.7	17.1
60-69	429	15.9 (3.1)	59.7	42.4	11.7	26.6	17.7
70-79	336	16.2 (3.1)	55.1	49.7	6.0	25.3	12.5
80-89	54	16.7 (3.1)	42.6	57.4	9.3	25.9	9.3
All	1867	15.5 (2.9)	55.4	36.3	16.0	22.7	18.0

Fig 1: Demographic details by age band of current CCAB normative dataset hosted on OSF

Summary

- Normative data from 1,867 participants in 40 tasks is publicly available on OSF, including high-precision, high-resolution speech metrics
- Additional normative data will be added as more subjects are run
- We further plan to add trial and/or response-level data, and Spanish-language results

CCAB

Our OSF page

More CCAB info

Contact

References

[1] Woods, D.L., et al. (2024). The California cognitive assessment battery (CCAB). *Frontiers in Human Neuroscience*, *17*, 1305529.

Contact us
mike_blank@neurobs.com for reprints
ccabresearch.com
neurobs.com

Visit us at booth 646!

Supported by NIA R44AG062076
NIA R44AG080951