AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

ISSN: 0894-4105

https://doi.org/10.1037/neu0001005

Concurrent Validity, Test–Retest Reliability, and Normative Properties of the Ignite App: A Cognitive Assessment for Frontotemporal Dementia

Rhian S. Convery[1], Kerala Adams-Carr[1], Jennifer M. Nicholas[2], Katrina M. Moore[3], Sophie Goldsmith[1], Martina Bocchetta[1, 4], Lucy L. Russell[1], and Jonathan D. Rohrer[1]

[1] Department of Neurodegenerative Diseases, Dementia Research Centre, Institute of Neurology, University College London
[2] Department of Medical Statistics, London School of Hygiene and Tropical Medicine
[3] Ionis Pharmaceuticals, Carlsbad, California, United States
[4] Department of Life Sciences, Centre for Cognitive and Clinical Neuroscience, Division of Psychology, College of Health, Medicine and Life Sciences, Brunel University London

Objective: Digital biomarkers can provide frequent, real-time monitoring of health-related behavior and could play an important role in the assessment of cognition in frontotemporal dementia (FTD). However, the validity and reliability of digital biomarkers as measures of cognitive function must first be determined. ***Method:*** The Ignite cognitive app contains iPad-based measures of executive function, social cognition, and other domains affected in FTD. Here we describe the normative properties of the Ignite tests, evaluate associations with gold-standard neuropsychological tests, and investigate test–retest reliability through two healthy control studies. Over 2,000 cognitively normal adults aged 20–80 years were recruited to complete the Ignite app remotely. A separate cohort of 98 healthy controls completed Ignite at two timepoints (7 days apart), a pen and paper neuropsychology battery, and a User Experience Questionnaire. ***Results:*** Significant associations were found between age and performance on several Ignite measures of processing speed ($r = 0.42$–0.56, $p < .001$) and executive function ($r = 0.43$–0.62, $p < .001$). With the exception of one test (Time Tap), the Ignite tests demonstrated moderate to excellent test–retest reliability (intraclass correlation coefficients [ICC] = 0.54–0.92) and significant correlations with their pen and paper counterparts ($r = 0.25$–0.72, $p < .05$). The majority of participants (>90%) rated the app favorably, stating it was enjoyable and easy to complete unsupervised. ***Conclusions:*** These findings offer early support for the validity of the Ignite tests suggesting they measure the intended cognitive processes, capture a stable picture of performance over time, and are well accepted in healthy controls. This work supports the feasibility of administering the app remotely and its potential utility as a cognitive tool in FTD; however, validation is ongoing, and further work is required before Ignite can be used as an endpoint in clinical trials.

Key Points
Question: How valid and reliable is the Ignite cognitive app in measuring key cognitive domains? ***Findings:*** Preliminary evidence suggests that the app offers a useful and consistent measure of cognitive performance, though its validation remains an ongoing process. ***Importance:*** Designed for remote use, the app allows individuals to complete assessments from home on an iPad, making it especially valuable for monitoring brain health over time, such as in clinical trials for frontotemporal dementia (FTD), a condition that impacts cognitive abilities. With its user-friendly design and positive reception from participants, Ignite shows promise as a research tool. ***Next Steps:*** Future studies will further explore its validity using updated versions of the app to ensure its effectiveness in detecting early cognitive impairments in individuals at risk of developing FTD.

Keywords: frontotemporal dementia, cognitive assessment, digital biomarkers

Supplemental materials: https://doi.org/10.1037/neu0001005.supp

Steven Paul Woods served as action editor.
Rhian S. Convery https://orcid.org/0000-0002-9477-1812
Kerala Adams-Carr https://orcid.org/0000-0002-8147-2610
Martina Bocchetta https://orcid.org/0000-0003-1814-5024
Lucy L. Russell https://orcid.org/0000-0001-5023-5893
Jonathan D. Rohrer https://orcid.org/0000-0002-6155-8417
This article was funded by the National Brain Appeal awarded to Rhian S. Convery. The authors thank Chris Frost and Amy MacDougall for additional statistical advice. Ionis has neither supported nor taken part in any aspect of this work.

Rhian S. Convery played a lead role in data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software,

continued

Frontotemporal dementia (FTD) is a clinically, genetically, and pathologically diverse neurodegenerative disorder. Clinically, FTD is characterized by behavioral change (behavioral variant FTD) or language impairment (primary progressive aphasia) and can overlap with amyotrophic lateral sclerosis and atypical parkinsonian disorders. The majority of FTD cases are sporadic; however, in approximately one third of individuals, symptoms are caused by an autosomal dominant genetic mutation in the chromosome 9 open reading frame 72 (*C9orf72*), microtubule-associated protein tau (*MAPT*), or progranulin (*GRN*) genes (Onyike & Diehl-Schmid, 2013; Rohrer et al., 2009). Evaluating first-degree relatives of individuals with genetic FTD provides an insight into the presymptomatic stage of the disease, where individuals do not have symptoms but have a 50% chance of carrying one of the mutations (Benussi et al., 2022).

There are currently no disease-modifying therapies available for FTD; however, clinical trials that target pathogenic mutations are underway, where therapies will likely be the most effective when administered early in the disease course (Tsai & Boxer, 2016). However, sensitive and validated biomarkers of disease onset and progression in the presymptomatic phase of FTD are lacking, particularly cognitive ones (Staffaroni et al., 2022). There are several limitations to the current approach of administering traditional pen and paper neuropsychology tasks, as they require participants to break from their normal routine, can lack reliability and consistency in administration and scoring, and can be time-consuming for patients. In addition, tasks appear to lack sensitivity to cognitive change in FTD prior to symptom onset (Rohrer et al., 2015). The ubiquitous use of technology may provide a solution for improving the current standard of cognitive assessments. Digital cognitive assessments can reduce interrater variability, time, and associated costs of testing and can be used as home monitoring tools, reducing patient burden and capturing cognition in a participant's natural environment. Furthermore, digital assessments allow for more frequent testing, could allow studies to collect more detailed data sets, and potentially yield more sensitive measures of cognition. Digital tools could therefore offer complementary value to traditional methods of assessing cognition if demonstrated to be reliable and valid measures. As such, several digital cognitive assessments have been developed for the early detection of mild cognitive impairment both for screening tools and clinical endpoints in Alzheimer's disease trials (Öhman et al., 2021; Onoda et al., 2013; Onoda & Yamaguchi, 2014; Papp et al., 2021; Possin et al., 2018; Weintraub et al., 2013). However, there are few digital cognitive assessments available that are specifically designed to measure cognitive impairment in FTD.

Ignite is a cognitive assessment app for the iPad, designed for FTD observational research and clinical trials. The app includes 12 unique tests measuring information processing speed, executive function, social cognition, semantic knowledge, arithmetic, and visuospatial skills (see Moore, Convery, Rohrer, 2022, for app protocol). Tests were included to tap into domains known to be affected in FTD, particularly presymptomatically, with the goal of improving sensitivity to detect cognitive impairment (Jiskoot et al., 2016, 2018; Moore, Convery, et al., 2022; Rohrer et al., 2015). Ignite is a self-assessment and is mainly comprised of computerized versions of standard neuropsychology tasks that can be completed in under 30 min. Before being tested in clinical cohorts to assess impairment, novel cognitive assessments need to provide appropriate validity and reliability estimates, including the development of new normative properties (Morrison et al., 2015; Ruggeri et al., 2016). Furthermore, demonstrating the feasibility of administering computerized cognitive assessments is equally important to ensure digital assessments are well accepted by the population. Here, we describe the normative properties and construct validity of the Ignite tests in a study of over 2,000 healthy controls (Study 1) and report concurrent validity with gold-standard neuropsychometry, test–retest reliability estimates, and feasibility data in a separate study of 98 healthy controls (Study 2).

Method

Study 1

Participants

Participants were recruited to complete Ignite remotely via media advertisements describing the study and explaining that healthy volunteers were required to complete the app. Previous research has demonstrated that the precision of normative data for cognitive tests improves as sample size increases, with confidence intervals narrowing as the sample grows, until reaching a plateau at approximately 200 cases (Crawford & Garthwaite, 2008). Based on this, a sample size of 200 individuals per decile was considered sufficient to power the normative data set, with the goal of achieving an equal distribution between sexes. Participants were encouraged to read the study information sheet on the first page of the app carefully to ensure they met the following eligibility criteria: they were healthy controls (i.e., did not have a significant neurological or psychiatric disorder), were aged between 20 and 80, owned an Apple iPad (any model), and were able to understand and comply with instructions in English. Participants were asked not to take part if they had a clinically significant medical condition that could affect their safety, preclude evaluation of response, or interfere with the compliance of study procedures, had a visual (noncorrected) or physical motor impairment that could interfere with their ability to use the iPad and/or complete the assessment, had learning difficulties or dyslexia, or if they were taking any medication that may impact their cognitive performance.

Procedure

Instructions were provided in the study advertisement describing how to download the Ignite app from the App Store onto the participant's personal devices. After reading the information sheet in the app and completing the consent form, participants were then

validation, visualization, and writing–original draft. Kerala Adams-Carr played a supporting role in investigation. Jennifer M. Nicholas played a supporting role in formal analysis. Katrina M. Moore played a supporting role in conceptualization whilst at University College London. Sophie Goldsmith played a supporting role in investigation. Martina Bocchetta played a supporting role in supervision. Lucy L. Russell played a supporting role in formal analysis and supervision. Jonathan D. Rohrer played a lead role in conceptualization and supervision and a supporting role in writing–review and editing.

Correspondence concerning this article should be addressed to Jonathan D. Rohrer, Department of Neurodegenerative Diseases, Dementia Research Centre, Institute of Neurology, University College London, London WC1N 3AR, United Kingdom. Email: j.rohrer@ucl.ac.uk

presented with a short form requiring them to enter basic demographic information, including age (years), education (years), sex (M/F), country of residence, and the first three letters of their city of birth (to help with the identification of duplicate attempts at the tests). No personally identifiable information was collected. Participants then completed 12 separate cognitive tasks in a predetermined order: Think Back Level 1 and Level 2, Sum Up, Colour Mix Levels 1–4, Face Match, Mind Reading, Swipe Out, Card Sort, Line Judge, Balloon Fair, Time Tap, Path Finder Levels 1 and 2, and Picture Pair (Table 1, Figure 1). Detailed instructions were presented at the beginning of each test accompanied by example videos demonstrating how the task should be completed. No feedback was provided to participants on their performance, and on the final page of the assessment, participants were required to select "Upload data," which sent the results to a secure server.

Data Preprocessing

Data from each participant were collated and analyzed in Stata/MP (Version 16.1). Participants were grouped into six age groups (20–29, 30–39, 40–49, 50–59, 60–69, 70–80 years) and four education groups (0–9, 10–12, 13–16, and ≥17 years). These education groupings were chosen to reflect standard levels of education in the United Kingdom. For each test in the assessment, the number of trials completed per person was calculated and averaged across each age group. For each test in the assessment, participant test data were excluded if the number of trials they had completed was more than 3 *SD*s below the mean for their age group. This criterion was applied to remove extreme outliers, ensuring there were a sufficient number of trials to analyze for each test and to remove participants that had not properly attempted the task.

Outcome Measures

For several tests, including Think Back, Colour Mix, Mind Reading, Face Match, Sum Up, Picture Pair, and Line Judge, three measures of speed (average reaction time across trials), accuracy (total correct), and a speed–accuracy trade-off (SAT) score = total correct/average reaction time were calculated for each test. Average reaction time was also calculated for the Swipe Out task along with the Flanker effect measure (average reaction time of incongruent trials—average reaction time of congruent trials). The total completion time (in seconds) was calculated for Path Finder Levels 1 and 2. The total number of correct categories achieved was the measure of interest for the Card Sort task, and the total amount of money won was calculated for the Balloon Fair test. Clock variance and absolute drift were computed from an autoregressive timing model for the Time Tap task (Henley et al., 2014), where clock variance represents a cognitive representation of time or "the internal clock," and absolute drift is the difference in the response interval from the first and last tap (see Supplemental Table S1 for a detailed description of outcome measure calculation).

Statistical Analysis

Demographic Associations. To examine the isolated effects of demographic predictors on performance, the partial correlations of age and education with each of the Ignite outcome measures were calculated, controlling for the remaining demographic variables (i.e., sex and age/education). Pearson's or Spearman's (if data were nonnormal) partial correlations were performed in R (RStudio Team, 2021). Linear regressions were used to analyze sex differences with males as the reference group, and age and education were included as covariates in the model. For outcomes that were not normally distributed, bootstrapping with 2,000 replications was used to calculate 95% confidence intervals. For the Card Sort task, as the outcome measure (number of correct categories) is categorical, a logistic regression was used to investigate if age, education, and sex predicted task performance. All regression analyses were performed in Stata/MP (Version 16.1).

Normative Properties. To produce normative scores by age deciles, education groups, and sex, adjusted means were output from linear regression estimates for each Ignite outcome measure. To generate a *Z*-score calculator from the normative data, multiple linear regressions were conducted for each Ignite outcome measure that adjusted for the demographic predictors of age, sex, and years in education, concurrently, individually and without covariates, resulting in five different models per measure. *Z*-scores were estimated by subtracting raw scores from the predicted mean(s) and then dividing this difference score by the standard deviation of the residuals (root-mean-squared error term). To ensure *Z*-scores were interpretable, outcome measures that were not normally distributed were transformed prior to analysis (see Supplemental Table S2). To investigate potential nonlinearity between age and cognitive performance, a quadratic term for age was added to each model (see Supplemental Table S3). From the data, we observed that with the exception of the Sum Up and Path Finder tests, the quadratic age terms were nonsignificant and did not substantially increase the explained variance of the models. For Sum Up and Path Finder, however, the results indicate a nonlinear relationship with age (Supplemental Figure S1), and we therefore used the quadratic term in the normative data estimation for these tests.

Construct Validity. To understand construct validity, an exploratory factor analysis was performed in Stata/MP (Version 16.1). We conducted two separate factor analyses in this study using a polychoric correlation matrix, chosen for its ability to include both continuous and categorical (i.e., Card Sort) variables. The primary analysis was carried out on all 43 outcome measures, both continuous and categorical (e.g., Card Sort), to examine interitem correlations. This analysis ensured that outcomes from the same tests were consistently loaded onto the same factors, with intercorrelations being notably high ($r > 0.90$). For the secondary factor analysis, we selected the outcomes with the highest loadings from each test, representing each task's key cognitive component, to avoid issues of multicollinearity and allow us to construct easily interpretable composite measures of cognition in future studies. We evaluated the scree plot of eigenvalues to determine the optimal number of factors, leading to the selection of a five-factor model (Supplemental Figure S2). Factor loadings were interpreted using the oblique promax rotation method and a minimum criterion of a primary loading factor of 0.3 or above was applied to each outcome measure (Ford et al., 1986).

Study 2

Participants

Participants were recruited to an observational study held at University College London (UCL). All participants gave full informed

Table 1
Ignite Tests With Corresponding Gold-Standard Pen and Paper Equivalents, the Cognitive (Subdomain) Measured, a Description of the Task, the Maximum Number of Possible Trials, and the Time Allowed to Complete Each Test (in Seconds)

Ignite test	Gold-standard test	Cognitive domain (subdomain)	Description	Number of trials	Time to complete(s)
Think Back (Level 1)	*N*-back (1 back)	Processing speed	Participants must indicate whether a shape presented exactly matches (including color) the shape that preceded it.	72	60
Think Back (Level 2)	*N*-back (2 back)	Executive function (working memory)	Participants must indicate whether a shape presented exactly matches (including color) the shape that came two before it.	72	60
Sum Up	Graded Difficulty Arithmetic Test	Arithmetic	Participants are required to select the correct answer (from four options) to different calculations including additions, subtractions, multiplications, and divisions.	48	60
Colour Mix (Level 1)	D-KEFS—color naming	Processing speed	Participants are required to tap the name of a color that matches the color of a circle.	50	30
Colour Mix (Level 2)	D-KEFS—word reading	Processing speed	Participants are required to match the name of a color that matches a color word.	50	30
Colour Mix (Level 3)	D-KEFS—color–word inhibition	Executive function (inhibitory control)	Participants are presented with a color word written in different colored ink. They are required to choose the name of the color that matches the color of the ink.	50	60
Colour Mix (Level 4)	D-KEFS—inhibition/switching	Executive function (inhibitory control/set-shifting)	Participants are presented with a color word written in different colored ink. They are required to complete the task as in Level 3, unless a black border appears around the word in which case, they should match the written word rather than the ink color.	50	60
Face Match	Ekman 60 Faces Test	Social cognition	Participants are required to tap the faces displaying a target emotion that match a word written at the top of the screen (e.g., Happy, Surprise, Anger, Fear, Sadness, Disgust).	30	60
Mind Reading	Reading the Mind in the Eyes task	Social cognition	Participants are required to tap the pair of eyes (from four options) that match a target emotion written at the top of the screen (e.g., reflective, contemplative).	20	90
Swipe Out	Eriksen Flanker task	Executive function (inhibitory control)	Participants are presented with arrows pointing in a given direction. They are required to swipe on the screen, in the direction of the central arrow. Flanking arrows are either facing in the same (congruent) or different (incongruent) direction.	40	60
Card Sort	Wisconsin Card Sorting task	Executive function (set-shifting)	Participants are presented with four cards, one in each corner of the screen, and one card in the middle. They are required to match the card in the middle to one of the four cards according to a secret rule (e.g., color, shape or number).	48	90
Line Judge	Benton Judgment of Line Orientation	Visuospatial skills	Participants are required to select two lines that match with the orientation of two target lines presented above.	30	90

(*table continues*)

Table 1 *(continued)*

Ignite test	Gold-standard test	Cognitive domain (subdomain)	Description	Number of trials	Time to complete(s)
Balloon Fair	Iowa Gambling Task	Executive function (decision making)	Participants are required to pump a balloon to win money. The more the balloon is inflated, the more money there is available to collect; however, if the balloon is pumped too much, it can burst.	50	90
Time Tap		Executive function (cognitive timing)	Participants are required to tap in time with a circle that pulses at regular intervals. After 30 s, the circle disappears, and the participant must continue tapping at the same tempo.	40	60
Path Finder (Level 1)	Trail Making Test part A	Processing speed	Participants are required to tap from one number to the next in sequential order.	19	90
Path Finder (Level 2)	Trail Making Test part B	Executive function (set-shifting)	Participants are required to alternate between tapping from numbers to letters in sequence (i.e., 1, A, 2, B).	19	90
Picture Pair	Modified Camel and Cactus test	Semantic knowledge	Participants are required to select an image (from four options) that best matches with a target image.	25	120

Note. D-KEFS = Delis–Kaplan Executive Function System.

consent at the beginning of the research visit. A sample size of 88 was calculated to provide 90% power for detecting a correlation of at least 0.5 (one-sided), assuming a moderate to strong correlation between assessments ($r \geq 0.7$). The same inclusion and exclusion criteria used in Study 1 were applied, with the exception that participants did not need to own an Apple iPad, and an additional requirement that they must not have completed the Ignite app before (i.e., as part of Study 1). Healthy controls were recruited through the online platform Join Dementia Research (https://www.joindementiaresearch.nihr.ac.uk/). Only individuals that met the study criteria received the details, and then these participants could choose to register to take part in the study. Participants did not receive compensation for taking part in the study.

Procedure

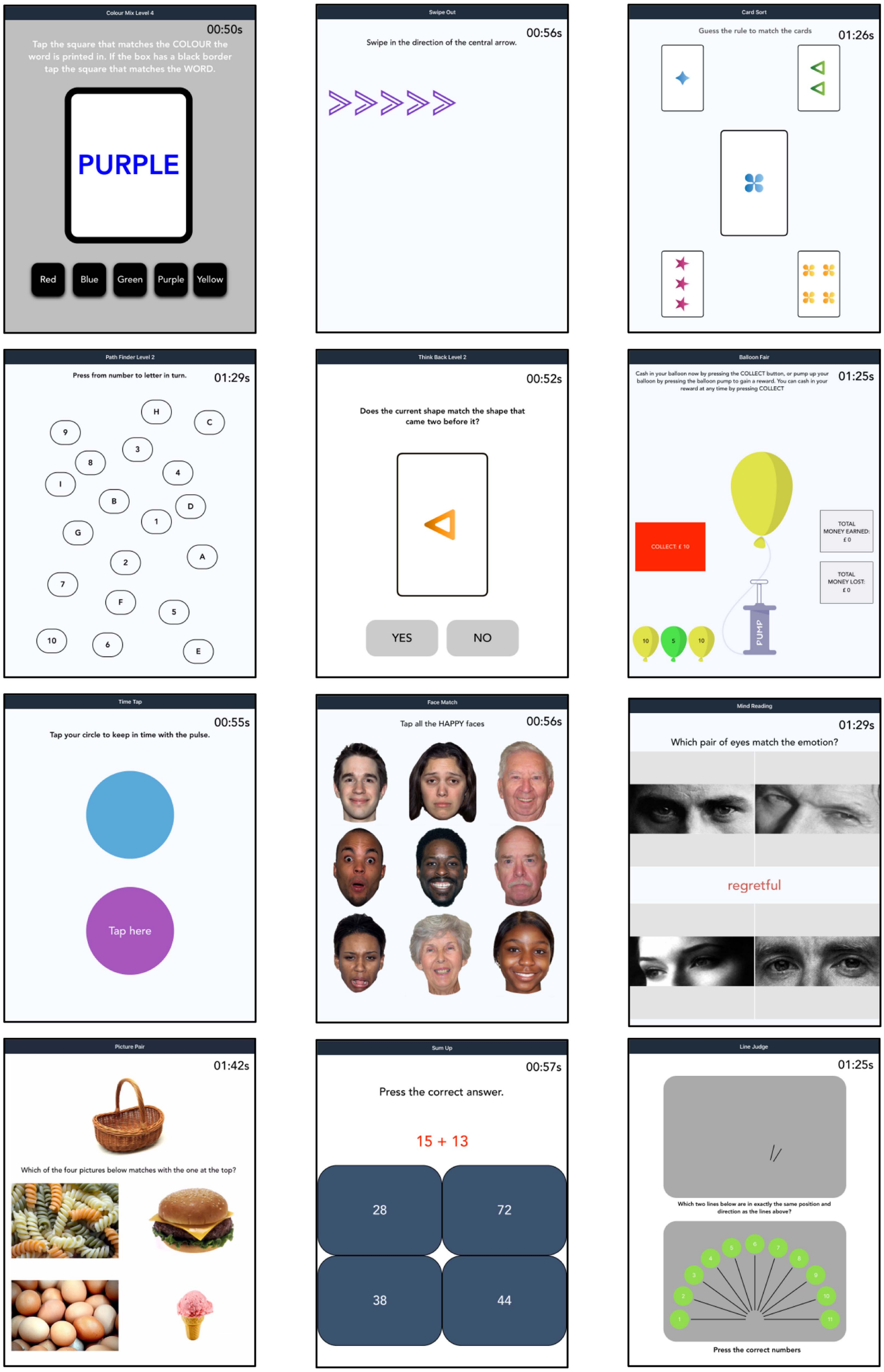
Participants attended two 1-hr research visits at UCL conducted 2 weeks apart. All participants completed the Ignite app and a neuropsychology battery containing gold-standard pen and paper versions of the Ignite tests. Participants were randomized 1:1 into two conditions, completing either the Ignite app or the neuropsychology battery at the first research visit. The pen and paper neuropsychology battery was administered in a quiet testing room and included 11 different tests (see Table 1). After the first research visit, participants were given a study iPad, with the Ignite app downloaded, and were asked to complete the assessment 1-week later. Therefore, participants completed Ignite at two timepoints (7 days apart), once remotely, and once during one of the research visits. At the end of the study, participants were invited to complete the Ignite User Experience Questionnaire, via email link to the Lime Survey platform (Version 2.28.34). The survey included 10 statements, concerning attitudes and experiences of completing Ignite, rated on a 5-point Likert scale (ranging from *strongly disagree* to *strongly agree*). To reduce response bias, statements were randomized so that 50% were of a positive attitude and 50% were negative.

Statistical Analysis

Concurrent Validity. A correlation analysis was conducted to determine the relationship between Ignite tests and the standard pen and paper neuropsychology tasks. Pearson correlations were computed for normal measures and Spearman correlations were used for outcome measures with a nonnormal distribution. A chi-squared test was conducted to assess the relationship between scores on the Card Sort task and the Wisconsin Card Sorting Task. The most comparable outcome measures were chosen to compare performance on Ignite tests with the neuropsychology battery. Where direct equivalents were not available, Ignite tests were correlated with other pen and paper tasks hypothesized to measure the same cognitive domains.

Test–Retest Reliability. A two-way mixed effects model for each Ignite outcome measures across the two timepoints was used to calculate the consistency of agreement intraclass correlation coefficients (CA-ICC; Shrout & Fleiss, 1979). Consistency of agreement ICCs were selected to allow for a difference in performance upon repeated assessment (e.g., learning effects; Koo & Li, 2016). Established cutoff criteria for CA-ICC values were used to determine the level of reliability for each outcome measure where poor $\leq$0.50, moderate = 0.50–0.75, good = 0.75–0.90, excellent $\geq$ 0.90 (Koo & Li, 2016; Portney & Watkins, 2000). The mixed model was also used to

Figure 1
Tasks in the Ignite Battery



Note. Tasks displayed: top left *Colour Mix*, top middle *Swipe Out*, top right *Card Sort*, second row left *Path Finder*, second row middle *Think Back*, second row right *Balloon Fair*, third row left *Time Tap*, third row middle *Face Match*, third row right *Mind Reading*, bottom left *Picture Pair*, bottom middle *Sum Up*, bottom right *Line Judge*. See the online article for the color version of this figure.

calculate the mean difference between scores and the two timepoints and the 95% upper and lower limits of agreement for the mean difference. Nonnormally distributed data were transformed prior to analysis. Bland–Altman plots were constructed to demonstrate agreement between Ignite scores and back-transformed for measures that were not normally distributed.

User Experience Questionnaire. The percentage of responses was calculated for each rating on the Likert scale. *strongly agree* and *agree* responses were collated into one single measure of "Agreement" and *strongly disagree* and *disagree* responses into one measure of "Disagreement" to improve the interpretability and provide an overall picture of the group attitude per question. To investigate potential differences in user experience by age, participants were split into two groups of younger (age <59 years) and older (age ≥60) adults, and chi-squared tests were used to assess differences in rating for each question.

Transparency and Openness

We report how we determined the required sample sizes for both studies, all data exclusions (if any), all manipulations, and all measures in the study. The data sets used and/or analyzed are available from the corresponding author on reasonable request. The normative calculator is available for use on the genetic frontotemporal dementia initiative website (https://www.genfi.org/ignite/). The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author. Data were analyzed using Stata/MP (Version 16.1) and RStudio Team (2021). This study's design and its analysis were not preregistered. The authors do not have any conflicts of interest to disclose with the Ignite app.

Results

Study 1

Participant Characteristics

A total of 2,043 people completed the Ignite app. There were nine different countries represented in the data set, with 95.3% of participants residing in the United Kingdom and 4.3% from the United States. The remaining 0.4% (N = 8) of participants were from countries where English was not the predominant language and were therefore excluded to limit the effects of language comprehension on task performance. Analysis of the data revealed 31 participants with identical demographic information to at least one other participant. These individuals were also excluded based on the possibility these could be duplicate participants, resulting in 39 healthy controls in total being excluded prior to analysis. Therefore, 2,004 participants were included in this study; see Table 2 for demographic information. Individual test data was removed based on the number of trials completed being lower than three standard deviations of the population average in 24 instances (Supplemental Table S4). Female participants accounted for 67.4% of the normative sample. The mean (standard deviation) age of the population was 55.2 (15.8), and the number of years in education was 16.1 (4.2).

Demographic Associations

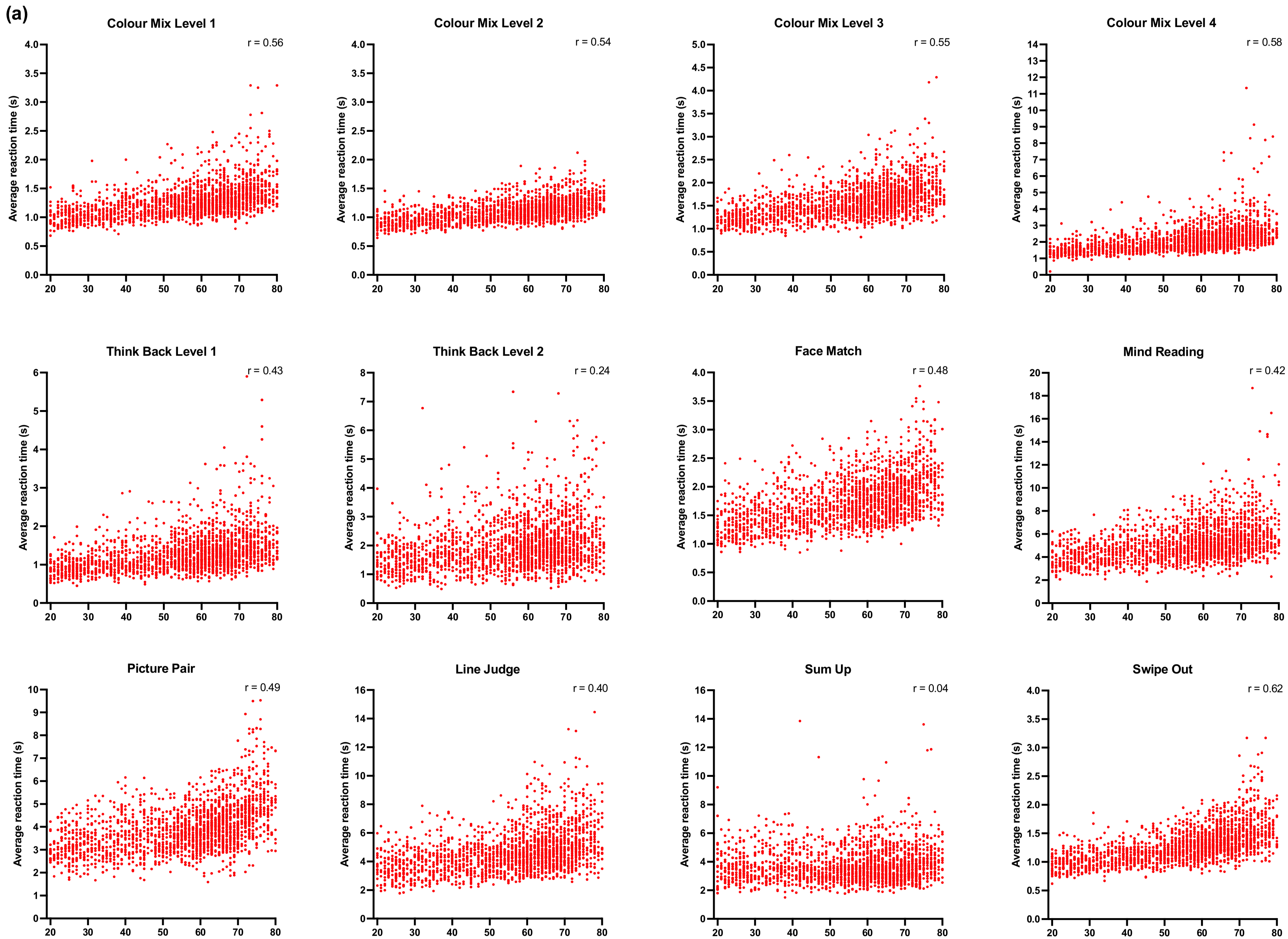
Significant associations between age and performance were seen for 38 out of the 43 Ignite outcome measures ($p < .01$). A positive correlation between age and average reaction time was observed across the tests (r = 0.24–0.62, $p < .001$), indicating a slowing of responses with age, accompanied by a decrease in accuracy in the total number of correct trials (r = −0.12 to −0.57, $p < .001$) and speed–accuracy trade-off scores (r = −0.25 to −0.58, $p < .001$), see Figure 2. A decline in performance was also seen with age and the total money earned on the Balloon Fair task (r = −0.36, $p < .001$) and the number of correct categories achieved in the Card Sort test (β = −0.06, $p < .001$). Only the Sum Up task did not display significant correlations with age. Significant sex differences were observed with females performing better on Colour Mix Levels 1–3 (β = 0.39–1.25, $p < .05$), Path Finder Level 2 (β = −2.30, $p < .001$), Face Match (β = 0.97–1.25, $p < .001$), Mind Reading (β = 0.55, $p < .001$), and Picture Pair (β = 0.43, $p < .05$). Male participants had significantly higher scores on Sum Up (β = 1.92, $p < .001$), Line Judge (β = 1.52, $p < .001$), Think Back Level 2 (β = 1.82, $p < .001$), and Balloon Fair tasks (β = 85.5, $p < .001$). Significant associations between education and performance were seen on several Ignite tests; however, effect sizes were small (see Table 3).

Table 2
Demographic Characteristics of the Healthy Control Participants in Each Study

Demographic	Study 1	Study 2	Study 2 (subsample who completed feasibility questionnaire)
Number of participants	2004	98	55
Age, years			
M (*SD*)	55.2 (15.8)	51.2 (17.3)	50.0 (17.8)
Age group (number of participants)			
20–29	200	16	10
30–39	203	15	9
40–49	233	14	8
50–59	394	17	8
60–69	587	18	10
70–80	387	18	10
Education, years			
M (*SD*)	16.1 (4.2)	17.8 (2.9)	17.9 (2.4)
Sex			
% Male	32.6	43.7	49.1

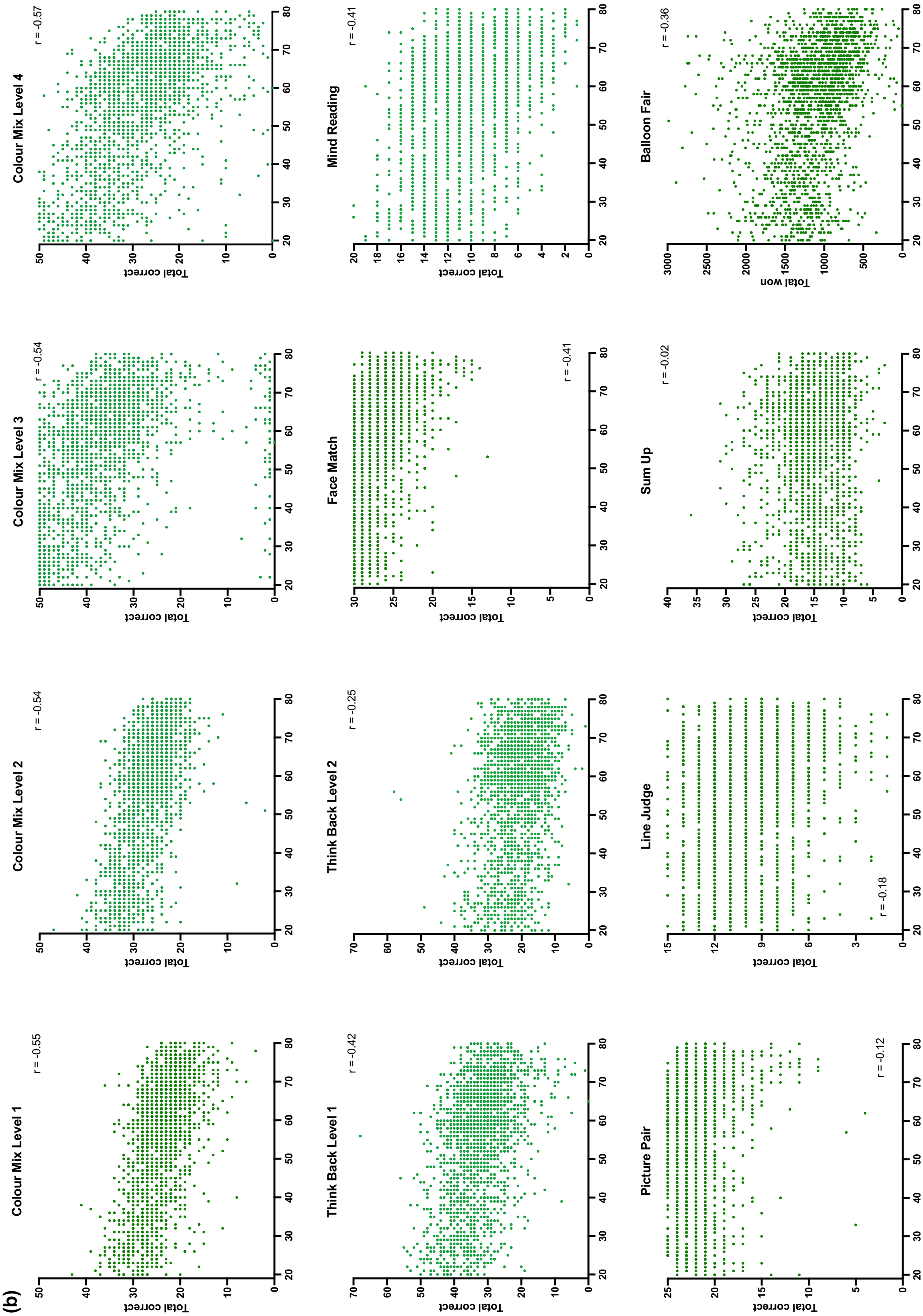
Figure 2

Relationships Between Different Ignite Outcome Measures and Age



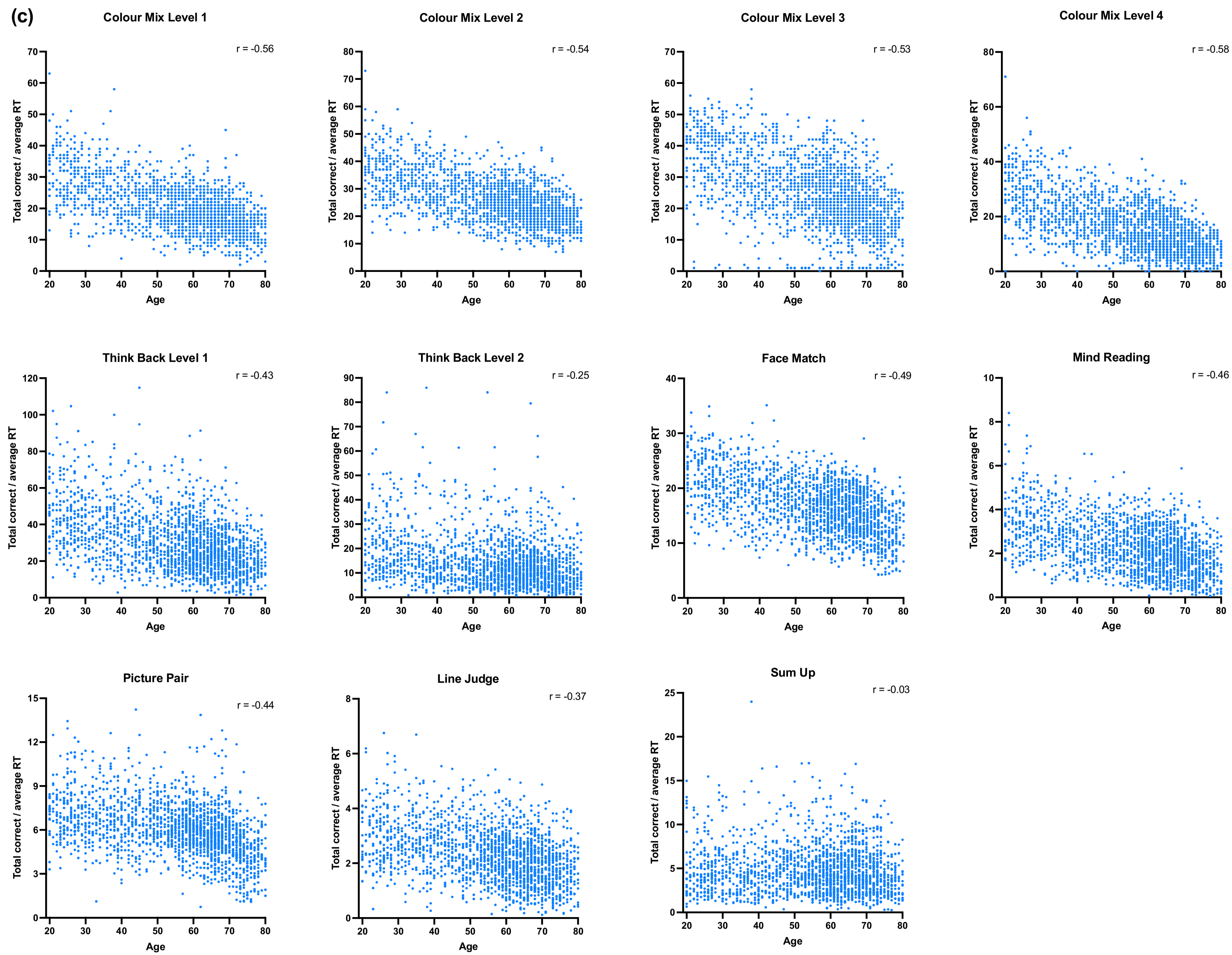
(*figure continues*)

Figure 2 *(continued)*



(figure continues)

Figure 2 *(continued)*



Note. Scatterplots displaying the relationship between (a) average reaction time (in seconds), (b) total correct, and (c) SAT scores with age on the Ignite tests. r = partial correlation coefficient. SAT = Speed–Accuracy Trade-Off score RT = reaction time. See the online article for the color version of this figure.

Table 3
Partial Correlations for Age and Education, and Linear Regression for Sex for Each Ignite Outcome Measure

Cognitive domain and subdomain	Ignite test (outcome measure)	Age (r)	Education (r)	Sex (β)
Processing speed	Path Finder Level 1			
	Time to complete (s)	0.50*	−0.07**	−0.50
	Colour Mix Level 1			
	Average reaction time (s)	0.56*	−0.02	−0.05*
	Total correct	−0.55*	0.02	0.77*
	SAT score	−0.56*	0.02	1.07*
	Color Mix Level 2			
	Average reaction time (s)	0.54*	−0.11	−0.02**
	Total correct	−0.54*	0.01	0.39**
	SAT score	−0.54*	0.01	0.60
	Think Back Level 1			
	Average reaction time (s)	0.43*	−0.04	−0.03
	Total correct	−0.42*	0.06**	0.20
	SAT score	−0.43*	0.05**	−0.29
Executive function				
Set-shifting	Path Finder Level 2			
	Time to complete (s)	0.43*	−0.10*	−2.30*
	Card Sort (β)			
	Number of categories completed	−0.06*	0.03**	−0.08
Inhibitory control	Color Mix Level 3			
	Average reaction time (s)	0.55*	−0.01	−0.56*
	Total correct	−0.54*	0.04	1.35*
	SAT score	−0.55*	0.04	1.25*
	Color Mix Level 4			
	Average reaction time (s)	0.58*	−0.04	−0.07
	Total correct	−0.57*	0.08*	0.59
	SAT score	−0.58*	0.07**	0.33
	Swipe Out			
	Flanker effect (ms)	0.06**	0.01	31.6**
	Average reaction time (s)	0.62*	−0.07**	0.03**
Decision making	Balloon Fair			
	Total money won	−0.36*	0.04	−85.5*
Working memory	Think Back Level 2			
	Average reaction time (s)	0.24*	0.00	0.12**
	Total correct	−0.25*	0.05**	−0.79**
	SAT score	−0.25*	0.02**	−1.82*
Cognitive timing	Time Tap			
	Clock variance (ms^2): paced	0.07**	0.00	448
	Clock variance (ms^2): self—paced	0.06**	−0.03	503
	Absolute drift (ms^2): paced	0.04	−0.02	−5.19
	Absolute drift (ms^2): self—paced	−0.06**	0.00	3.95
Social cognition	Mind Reading			
	Average reaction time (s)	0.42*	0.05**	−0.04
	Total correct	−0.41*	0.00	0.55*
	SAT score	−0.46*	−0.04	0.12**
	Face Match			
	Average reaction time (s)	0.48*	0.03	−0.09*
	Total correct	−0.41*	0.00	0.80*
	SAT score	−0.31*	0.02	0.97*
Semantic knowledge	Picture Pair			
	Average reaction time (s)	0.49*	0.03	−0.11**
	Total correct	−0.12*	0.06**	0.43**
	SAT score	−0.44*	0.00	0.11
Visuospatial skills	Line Judge			
	Average reaction time (s)	0.40*	−0.04	0.46*
	Total correct	−0.18*	0.12*	−1.52*
	SAT score	−0.37*	0.10*	−0.61*
Calculation	Sum Up			
	Average reaction time (s)	0.04	−0.08**	0.38*
	Total correct	−0.02	0.08*	−1.92*
	SAT score	−0.03	0.08*	−1.19*

Note. r = correlation coefficient, β = linear regression coefficient with males as the reference group, SAT = Speed–Accuracy Trade-Off score.
* $p < .001$. ** $p < .05$.

Normative Data

Equations for calculating the Z-scores for each outcome measure were used to generate a normative calculator for raw scores. Percentile ranks were also calculated from the normal distribution of each Z-score. The calculator was constructed to provide five corresponding estimated Z-scores, per Ignite outcome measure, based on predictions from each linear regression model. Therefore, an individual's raw Ignite scores and demographic information can be input into the spreadsheet, and Z-scores and percentile ranks are subsequently generated for each adjustment. The normative calculator is now available on the Genetic Frontotemporal dementia Initiative website (https://www.genfi.org/ignite/). Age- and education-grouped normative values, as well as differences in sex, for each Ignite outcome measure are reported in Supplemental Table S5.

Construct Validity

The five-factor solution from the factor analysis explained 86.5% of the variance in the data. Swipe Out and Colour Mix Levels 1, 2, 3, and 4 were contained in Factor 1, while Face Match, Mind Reading, and Picture Pair loaded onto Factor 2 (see Tables 4 and 5). Outcome measures from Line Judge and Sum Up loaded together in Factor 3, and both Think Back Levels 1 and 2 loaded with Factor 4. Cross-loading was observed for Colour Mix Level 3 and Level with measures also loading with Card Sort on Factor 5. Factors were subsequently grouped with the following labels: (a) Processing speed/inhibitory control, (b) Social/semantic processing, (c) Visuospatial/arithmetic processing, (d) Working memory, and (e) Set-shifting (see Figure 3). Path Finder Levels 1 and 2, Balloon Fair, and Time Tap did not meet the minimum criteria (>0.3) of the primary loading factor and therefore did not group under a simple factor structure.

Study 2

Participant Characteristics

A total of 98 healthy controls (43.7% male) with a M (SD) age of 51.2 (17.3) years, and number of years in education of 17.8 (2.9), were recruited (Table 2). A subset of participants (N = 55) completed the Ignite User Experience questionnaire. This subsample had a M (SD) age of 50.0 (17.8) years and number of years in education of 17.9 (2.4), and 47.3% of the sample were male.

Concurrent Validity

The majority of the Ignite tests significantly correlated with their pen and paper counterparts (r = 0.25–0.72, p < .05), and other neuropsychology tests measuring the same cognitive domains where direct comparisons were not available (r = 0.31–0.73, p < .05; Table 6). The strongest correlations were observed for Sum Up (r = 0.72) and Colour Mix Level 3 (r = 0.71) tasks with their pen and paper equivalents, while a weaker relationship was shown for Colour Mix Level 2 (r = 0.25). Only the Face Match, Card Sort, and Time Tap tasks did not significantly correlate with corresponding neuropsychology tests.

Test–Retest Reliability

With the exception of the Time Tap task (ICC's = −0.12 to 0.12), the Ignite tests demonstrated moderate to excellent test–retest reliability estimates (ICC's = 0.54–0.92; Table 7). High levels of agreement were also observed in the majority of Ignite outcome measures as demonstrated by Bland–Altman plots (see Supplemental Figure S3).

User Experience Questionnaire

No significant differences in responses were found between younger and older participants on any of the items. Using a 5-point

Table 4
Rotated Factor Loadings for the Five-Factor Model of Ignite Outcome Measures

Test	Outcome measure	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Path Finder Level 1	Completion time					
Path Finder Level 2	Completion time					
Swipe Out	Reaction time	−0.462				
Colour Mix Level 1	Total correct	0.705				
Colour Mix Level 2	Total correct	0.744				
Colour Mix Level 3	SAT score	0.462				0.342
Colour Mix Level 4	SAT score	0.369				0.368
Card Sort	Categories complete					0.456
Balloon Fair	Total score					
Face Match	SAT score	0.310	0.613			
Mind Reading	SAT score		0.681			
Picture Pair	SAT score		0.550			
Sum Up	SAT score			0.327		
Line Judge	SAT score			0.546		
Think Back Level 1	Total correct				0.572	
Think Back Level 2	Total correct				0.600	
Time Tap—Paced	Absolute drift					
Time Tap—Self paced	Absolute drift					

Note. Loadings greater than 0.3 are shown. SAT = Speed–Accuracy Trade-Off.

Table 5
Between Factor Correlations for Rotated Factor Matrix

Factor	1	2	3	4	5
1	0.88	0.85	0.79	0.77	0.60
2	0.02	−0.44	0.35	0.43	−0.12
3	−0.38	0.12	0.13	0.29	−0.23
4	0.22	0.02	−0.45	0.34	−0.47
5	−0.13	−0.01	−0.16	0.15	0.60

scale, all participants (100%) reported (neutral or above) that the test example videos were helpful, they found the tests interesting, and they enjoyed completing the assessment. Additionally, almost all participants (neutral or above) agreed that the task instructions were easy to understand (96.4%), the tests were not boring or repetitive (94.5%), the assessment was not difficult to complete from home (98.2%), and the iPad was not difficult to use (98.2%; Figure 4).

Discussion

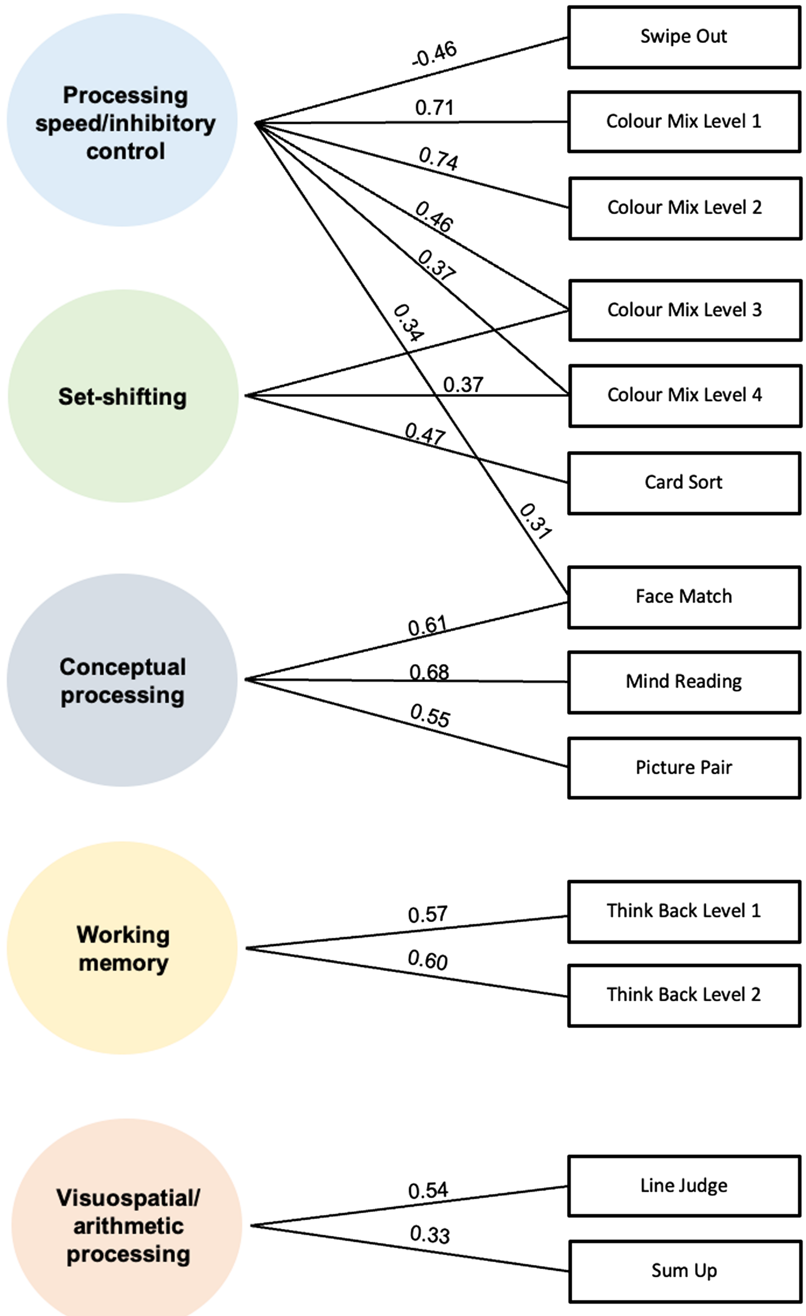
The Ignite app was tested in a population of 2,004 healthy controls, generating the largest normative data set to date for a computerized cognitive assessment of FTD. The results from Study 1 suggested the Ignite tests capture well-established trajectories of age-related cognitive decline seen in traditional pen and paper versions of tasks on measures of processing speed (McDowd & Shaw, 2000; Rasmusson et al., 1998; Tombaugh, 2004), executive function (Dempster, 1992; Mayr et al., 2001), social cognition (Charlton et al., 2009; Kemp et al., 2012; Kessels et al., 2014; Maylor et al., 2002), semantic knowledge (Hoffman, 2019; Wu & Hoffman, 2022), and visuospatial skills (Eslinger & Benton, 1983). Notably, there was a strong decline in performance with age on the Colour Mix and Path Finder tasks, corroborating studies of traditional versions of these tests that show increasing completion times with age on the Stroop task (Milham et al., 2002) and Trail Making Tests, respectively (Rasmusson et al., 1998; Tombaugh, 2004). However, the regression analyses conducted during the generation of normative data revealed that the relationship between age and completion time on the Path Finder tests was not linear. In fact, performance on these tasks remained relatively stable through midlife (in the 40s and 50s), after which it declined more rapidly with age. Additionally, significant correlations were not found between age and performance on the Sum Up test, which can likely be attributed to the nonlinear relationship revealed from the regression model. The fitted model suggests the cumulative acquisition of arithmetic knowledge over time, as arithmetic is a learned cognitive function that tends to improve with experience and education during midlife, before declining with age-related cognitive changes (Hartshorne & Germine, 2015).

Several significant differences in sex were observed for the Ignite tests, with female participants scoring higher on tasks assessing social cognition, replicating well-established findings that sex influences performance on emotion processing (Paletta et al., 2022), and empathy tests (Di Tella et al., 2020). Female participants also performed higher on Colour Mix Levels 1–3, while male subjects performed better on Think Back Level 2, Sum Up, Line Judge, and Balloon Fair. This corroborates the literature describing sex differences in cognitive function that suggest females score higher on tasks assessing cognitive flexibility and inhibitory control, while men tend to perform better on working memory, visuospatial processing and decision-making tasks (Lejuez et al., 2002; Lynn & Irwing, 2008; Mackintosh, 1996; Pawlowski et al., 2008; Singh et al., 2022). Several significant associations with education were observed for the Ignite tests; however, this was likely driven by the large sample size as the associations were small in magnitude. A regression-based normative calculator was subsequently developed that will be useful for assessing cognitive performance at the individual level in future studies. Utilizing adjusted *Z*-scores to assess performance on an individual basis could increase the sensitivity of the Ignite tests in detecting subtle deficits in presymptomatic FTD mutation carriers.

An exploratory factor analysis was conducted to identify the underlying factor structure of Ignite, as this approach is well-suited for evaluating the psychometric properties of a novel tool in its early stages of validation. Following an initial factor analysis to ensure measures from the same test loaded together, the highest loading measures from each task were retained for subsequent analysis. This allows for a streamlined representation of each task's key cognitive component, which could be used to create more interpretable and practical composite scores of cognition in future studies. The Ignite tasks were found to load onto factors that mostly align with the expected constructs. Specifically, all levels of the Colour Mix task loaded together on Factor 1, which likely suggests a shared reliance on processing speed and executive function. The additional loading of Swipe Out, along with Colour Mix Levels 3 and 4, implies that these tasks also tap into inhibitory control mechanisms, a finding that has been consistently demonstrated in studies utilizing traditional versions of these tests (Miyake et al., 2000). The strong correlations between Factor 1 and the other factors suggest that processing speed and inhibitory control may represent a general construct that explains a significant portion of the variance explained by all of the cognitive tasks, reflecting the interconnected nature of these cognitive domains. Conversely, the relatively weaker correlations between the remaining factors indicate that these may represent more specific cognitive functions. For example, the simultaneous loading of Colour Mix Levels 3 and 4 with the Card Sort task on a separate factor suggests that these tasks engage set-shifting abilities, indicating that higher order executive functions have been distinguished into distinct subcomponents. The fact that both Think Back levels loaded together reinforces the clear delineation of executive abilities as these tasks are known to be associated with working memory (Kane et al., 2007).

Tests of social cognition (Face Match and Mind Reading) and semantic knowledge (Picture Pair) unexpectedly loaded together. However, one theoretical perspective has grouped these processes together within the controlled semantic cognition framework (Binney & Ramsey, 2020; Ralph et al., 2017) which represents a conceptual knowledge base of the meaning of words, objects, and people, including person identification, empathy, and emotion recognition (Binney & Ramsey, 2020; Chiou et al., 2018). Finally, the coloading of the Line Judge and Sum Up tasks may reflect the fact that these tests likely engage multiple processes, particularly those related to posterior cortical functioning. This is consistent with the intended purpose of these tasks, as individuals with *C9orf72* and *GRN* mutations often exhibit early posterior cortical involvement. The observed factor structure supports this inclusion, aligning with neuropsychological evidence of the posterior

Figure 3
Factor Structure of the Ignite Tests



Note. The circles are labeled according to the cognitive domain each factor represents. The values denote the highest factor loading for an outcome measure in each test. This figure presents a simplified schematic of the results from the exploratory factor analysis, focusing on the highest factor loadings (>0.3) for each of the tests, omitting factor correlations and non-significant loadings. The figure is intended for illustrative purposes only and does not follow standard Reticular Action Model (RAM) notation. See the online article for the color version of this figure.

cortex's role in complex visual and numerical processing (Seeley et al., 2008). It is likely that the Path Finder and Balloon Fair tests did not load onto a single factor because they tap into a wider range of cognitive abilities, which may not be as closely related to the more general cognitive functions captured by other tasks (e.g., processing speed or executive function). Taken together, these results suggest that several of the tasks measured in the Ignite app tap into a range of cognitive abilities, with a general processing

Table 6
Correlations Between Ignite and Pen and Paper Neuropsychology Tasks

Ignite test (outcome measure)	Pen and paper test (outcome measure)	Correlation coefficient	*p*
Path Finder Level 1 (completion time)	Trail Making Test part A (completion time)	0.55	<.001
Path Finder Level 2 (completion time)	Trail Making Test part B (completion time)	0.52	<.001
Colour Mix Level 1 (average RT)	D-KEFS—color naming (completion time)	0.50	<.001
Colour Mix Level 2 (average RT)	D-KEFS—word reading (completion time)	0.25	.013
Colour Mix Level 3 (average RT)	D-KEFS—color-word inhibition (completion time)	0.71	<.001
Picture Pair (total correct)	Modified Camel and Cactus test (total correct)	0.42	<.001
Line Judge (total correct)	Benton Judgment of Line Orientation (total correct)	0.58	.000
Face Match (total correct)	Ekman 60 Faces Test (total correct)	0.17	.098
Mind Reading (total correct)	Reading the Mind in the Eyes task (total correct)	0.38	<.001
Sum Up (total correct)	Graded Difficulty Arithmetic Test (total correct)	0.72	<.001
Balloon Fair (total won)	Iowa Gambling Task (net total)	0.31	.004
Swipe Out (average reaction time)	Eriksen Flanker task (average reaction time)	0.52	<.001
Think Back Level 2 (total correct)	*N*-back (2 back; total correct)	0.46	.002
Tests without standard equivalent			
Colour Mix Level 4 (average RT)	D-KEFS—color-word inhibition (completion time)	0.73	<.001
Think Back Level 1 (average RT)	D-KEFS—color naming (completion time)	0.44	<.001
Think Back Level 1 (average RT)	Trail Making Test part A (completion time)	0.52	<.001
Time Tap–Paced (clock variance)	D-KEFS—color-word inhibition test (completion time)	0.13	.211
Time Tap–Paced (absolute drift)	Trail Making Test part B (completion time)	0.15	.157
Time Tap–Self-Paced (clock variance)	D-KEFS—color-word inhibition (completion time)	0.08	.471
Time Tap–Self-Paced (absolute drift)	Trail Making Test part B (completion time)	0.02	.840

Note. D-KEFS = Delis–Kaplan Executive Function System; RT = reaction time.

speed and inhibitory control factor likely acting as a core underpinning construct.

The results from Study 2 demonstrated that most of the Ignite tests exhibit good concordance with their corresponding pen and paper counterparts, supporting concurrent validity. Strong correlations were observed for Colour Mix Level 3 with Delis–Kaplan Executive Function System (D-KEFS) Ink Naming and Sum Up with Graded Difficulty Arithmetic, indicating the Ignite versions of these tasks track closely with the traditional versions. Additionally, Ignite tests without direct pen and paper comparisons correlated with other tests that measure the same hypothesized cognitive domains. The only tests that were not significantly associated with traditional measures were the Face Match, Card Sort, and Time Tap tasks. The lack of association between Face Match/Ekman Faces Test and Card Sort/Wisconsin Card Sorting Task can likely be explained by the simplicity of the pen and paper versions of these tasks and the resulting ceiling effects observed. Ceiling effects restrict the range of scores and result in low correlation coefficients. Furthermore, a high degree of variability was observed in the Time Tap data in both outcome measures, suggesting perhaps that cognitive timing is not a stable trait and could be highly influenced by other factors.

The Ignite tests displayed moderate to excellent test–retest reliability overall. Good test–retest coefficients >0.75 (Portney & Watkins, 2000) were obtained for all levels of the Colour Mix and Think Back tasks, as well as Sum Up, Line Judge, Face Match average reaction time (RT) and Speed–Accuracy Trade-Off scores, Picture Pair average RT scores, and Swipe Out average RT scores. Thus, findings are also consistent with previous studies demonstrating good reliability in pen and paper versions of these tasks (Adams et al., 2015; Franzen, 2000; Goldstein & Watson, 1989; Iverson, 2001; Lemay et al., 2004; Paap & Sawi, 2016; Salinsky et al., 2001; Sanders et al., 2018; White et al., 2018). Ignite reliability estimates for Think Back levels of working memory were higher than previously reported from 1-back (Lowe & Rabbitt, 1998; White et al., 2018) and 2-back tasks (White et al., 2018). Consistent with prior research studies using pen and paper (Barr, 2003) and computerized Trail Making Tests (Bracken et al., 2018; Morrison et al., 2015), the Ignite Path Finder levels displayed moderate reliability. Low test–retest reliability scores were only seen for the flanker effect measure on the Swipe Out task and all outcome measures produced from the Time Tap test. Consistent with the results from this study, other research has shown that reliability is low for the flanker effect but good for average reaction times on traditional Flanker tasks, both in healthy controls (Paap & Sawi, 2016; White et al., 2018) and in patients with dementia (Sanders et al., 2018). It is unclear why the Time Tap task displayed such low test–retest reliability scores, but the answer likely lies in the high variability seen in the data at both baseline and follow-up assessments. Nevertheless, the low estimates observed for these measures raise the longstanding issue as to whether tests with inadequate reliability should be excluded from future assessments.

The results of the Ignite User Experience Questionnaire demonstrate that healthy controls rate the app favorably overall. The majority of participants reported that the image quality of the tests was good, the videos were helpful, and the instructions were easy to follow. In addition, healthy controls agreed that the app was easy to complete remotely from home, and the iPad was not difficult to use. Demonstrating the feasibility of administering novel computerized assessments is equally as important as proving validity and reliability. Therefore, this data provides supporting evidence of the acceptability of the app among healthy adults (including those of older ages) and the feasibility of implementing Ignite as a cognitive test in the wider population through remote data collection studies.

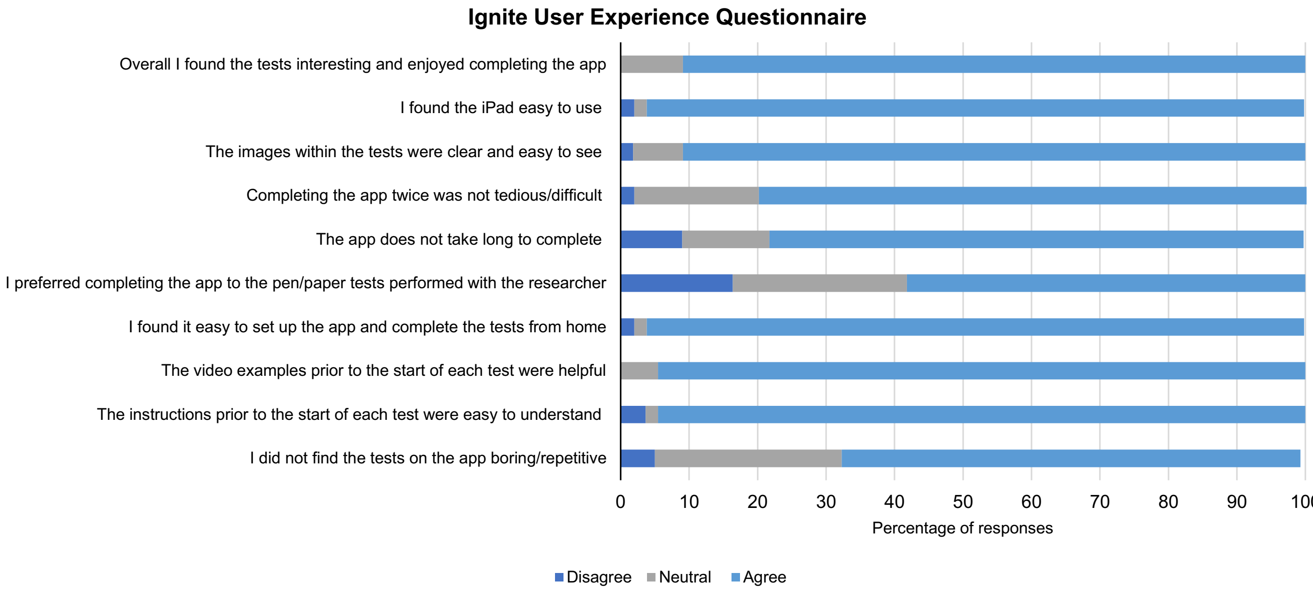
Several limitations to this work must be noted. First, to collect normative data, participants were recruited through convenience sampling and self-identified as not having a neurological or neuropsychological disorder. While those recruited through the join dementia research platform were prefiltered to exclude individuals

Table 7
Test–Retest Reliability Data for Ignite Outcome Measures

Test name	Outcome measure	Timepoint 1	Timepoint 2	Difference	CA-ICC	95% CI *LL*	95% CI *UL*	Lower agreement	Upper agreement
Sum Up	Total correct	15.5 (5.06)	17.1 (5.18)	1.62 (2.32)	0.90	0.85	0.93	−3.01	6.26
	Average RT	3.74 (1.29)	3.34 (1.01)	−0.40 (0.67)	0.89	0.84	0.93	−0.38	0.18
	SAT score	4.99 (3.10)	5.99 (3.48)	1.00 (1.36)	0.88	0.83	0.92	−0.39	0.81
Colour Mix Level 1	Total correct	24.3 (5.05)	25.9 (5.35)	1.63 (3.19)	0.81	0.73	0.87	−4.75	8.01
	Average RT	1.27 (0.32)	1.18 (0.29)	−0.09 (0.22)	0.82	0.74	0.88	−0.15	0.27
	SAT score	20.9 (8.29)	23.9 (9.57)	3.02 (5.20)	0.82	0.74	0.88	−0.83	1.46
Colour Mix Level 2	Total correct	27.5 (4.50)	28.5 (5.12)	0.96 (2.70)	0.84	0.77	0.89	−4.44	6.36
	Average RT	1.09 (0.19)	1.06 (0.20)	−0.03 (0.12)	0.84	0.77	0.89	−0.15	0.22
	SAT score	26.5 (8.52)	28.6 (9.99)	2.07 (5.03)	0.83	0.73	0.90	−7.99	12.1
Colour Mix Level 3	Total correct	39.1 (8.22)	41.4 (7.50)	2.23 (3.62)	0.89	0.85	0.93	−5.02	9.48
	Average RT	1.54 (0.38)	1.43 (0.36)	−0.11 (0.17)	0.92	0.89	0.95	−0.07	0.17
	SAT score	28.0 (11.2)	31.5 (11.6)	3.56 (4.47)	0.92	0.89	0.85	−5.39	12.5
Colour Mix Level 4	Total correct	31.3 (9.18)	33.7 (9.61)	2.41 (5.15)	0.85	0.78	0.90	−7.89	12.7
	Average RT	1.95 (0.63)	1.82 (0.59)	−0.13 (0.36)	0.87	0.81	0.91	−0.37	0.23
	SAT score	18.8 (9.80)	21.7 (11.3)	2.94 (5.31)	0.87	0.81	0.91	−0.92	1.55
Think Back Level 1	Total correct	31.5 (8.55)	37.2 (8.24)	5.65 (5.68)	0.77	0.68	0.84	−5.72	17.0
	Average RT	1.36 (0.60)	1.09 (0.47)	−0.27 (0.35)	0.73	0.62	0.81	−0.25	0.63
	SAT score	28.7 (16.0)	40.9 (19.9)	12.2 (13.1)	0.77	0.67	0.84	−1.09	3.19
Think Back Level 2	Total correct	20.4 (6.38)	24.1 (6.75)	3.68 (4.69)	0.75	0.64	0.82	−5.69	13.1
	Average RT	2.17 (0.89)	1.75 (0.76)	−0.42 (0.60)	0.76	0.66	0.83	−0.75	0.32
	SAT score	12.2 (9.13)	17.4 (11.0)	5.20 (6.70)	0.77	0.68	0.84	−0.97	2.33
Path Finder Level 1	Completion time	14.9 (6.21)	13.2 (4.33)	−1.67 (4.40)	0.72	0.61	0.80	−0.02	0.04
Path Finder Level 2	Completion time	33.2 (18.2)	28.6 (15.9)	−4.57 (13.7)	0.68	0.56	0.78	−0.88	0.59
Face Match	Total correct	27.3 (2.28)	27.9 (1.86)	0.62 (1.94)	0.57	0.42	0.69	−3.26	4.50
	Average RT	1.74 (0.38)	1.64 (0.39)	−0.11 (0.27)	0.75	0.65	0.83	−0.65	0.43
	SAT score	16.7 (4.79)	18.3 (5.54)	1.68 (3.60)	0.75	0.65	0.83	−0.66	1.05
Mind Reading	Total correct	11.5 (3.64)	13.0 (2.96)	1.47 (2.73)	0.66	0.54	0.76	−3.98	6.92
	Average RT	4.96 (1.43)	4.19 (1.06)	−0.77 (1.07)	0.73	0.62	0.81	−0.55	0.23
	SAT score	2.67 (1.35)	3.40 (1.38)	0.74 (0.86)	0.75	0.65	0.83	−0.34	0.81
Picture Pair	Total correct	22.2 (2.61)	22.9 (1.83)	0.73 (1.93)	0.63	0.50	0.74	−3.13	4.59
	Average RT	3.80 (1.27)	2.93 (0.99)	−0.86 (0.75)	0.84	0.77	0.89	−0.62	0.11
	SAT score	6.64 (2.61)	8.68 (2.77)	2.04 (1.40)	0.85	0.79	0.90	−0.16	0.92
Line Judge	Total correct	10.3 (2.32)	10.3 (2.62)	0.01 (1.69)	0.77	0.67	0.84	−3.37	3.39
	Average RT	4.59 (1.33)	3.99 (1.16)	−0.60 (0.77)	0.83	0.76	0.89	−0.46	0.18
	SAT score	2.45 (0.96)	2.81 (1.13)	0.36 (0.65)	0.82	0.74	0.87	−0.30	0.50
Balloon Fair	Total earned	1,145.9 (456.1)	1,369.8 (516.8)	223.9 (399.3)	0.68	0.56	0.77	−7.80	14.2
Swipe Out	Flanker effect (ms)	265.0 (258.1)	204.1 (168.5)	−60.9 (230.0)	0.44	0.27	0.59	−520.9	399.1
	Average RT	1.28 (0.32)	1.13 (0.27)	−0.15 (0.17)	0.89	0.85	0.93	−0.08	0.28
Card Sort	Number of categories completed	2.37 (1.34)	2.86 (1.18)	0.49 (1.03)	0.67	0.54	0.76	−1.57	2.55
Time Tap–Paced	Clock variance (ms2)	−1,190.2 (42,156.7)	−4,916.3 (27,496.2)	−3,726.0 (53,158.8)	−0.12	−0.31	0.09	−1,129.0	118,705.0
	Absolute drift (ms)	95.7 (92.3)	86.8 (89.3)	−8.90 (131.9)	0.12	−0.09	0.31	−3.03	2.98
Time Tap–Self-Paced	Clock variance (ms^2)	6,300.1 (21,121.8)	5,617.6 (18,194.75)	−682.5 (28,837.1)	−0.07	−0.27	0.14	−5.80	5.77
	Absolute drift (ms)	112.5 (97.9)	128.9 (101.9)	16.4 (139.9)	0.02	−0.18	0.22	−12.0	13.9

Note. Data represents mean (and standard deviation) scores on Ignite tests for Timepoint 1 and Timepoint 2 (7 days later) as well as the mean (standard deviation) for the difference between timepoints. Intraclass correlation coefficients measuring consistency of agreement (CA-ICC) between scores, the 95% confidence intervals (CI), and the upper and lower 95% limits of agreement are shown. *LL* = lower limit; *UL* = upper limit; SAT = Speed–Accuracy Trade-Off; RT = reaction time.

Figure 4
Ignite User Experience Questionnaire Results



Note. The bar chart displays the percentage of healthy controls that Agree, Disagree, or Feel Neutral to each statement. Statements of a negative attitude were inversed for interpretability. See the online article for the color version of this figure.

with diagnosed conditions, this approach did not include rigorous screening for undiagnosed cognitive deficits. The inclusion of individuals with potential undiagnosed conditions could introduce variability in the normative data, which may slightly influence thresholds or reduce sensitivity when interpreting individual performance. However, this variability reflects the natural diversity of the general population we aim to represent, which is important for ensuring Ignite's applicability in real-world contexts.

Second, ethnocultural and socioeconomic status data was not collected. Participants in Study 1 mainly resided in the United Kingdom, and only participants who spoke English and owned an Apple iPad were able to participate, limiting the generalisability of these findings. Third, we only collected data from individuals aged 20–80. Consequently, the associations between age and cognitive performance appear to be predominantly linear, as the sample excludes younger age groups where these cognitive abilities are acquired. Many cognitive functions likely peak in early adulthood, around the 20s, following a period of rapid development during childhood and adolescence. By not including individuals under 20, we miss the earlier portion of the "U-shaped" curve, where cognitive performance gradually improves before reaching its peak. This limitation restricts our ability to capture the full trajectory of cognitive development and its potential nonlinear relationship with age across the lifespan. Future studies of Ignite will incorporate young adults and teenagers in order to fully understand these developmental patterns.

Finally, the Ignite app does not currently include a measure of effort-testing or participant engagement. As a result, we cannot assess the performance validity of the normative sample. The inclusion of effort and engagement measures is particularly critical for self-administered cognitive tests, where engagement levels can vary widely due to distraction, fatigue, or lack of motivation. This raises concerns over the interpretability of results and highlights the importance of performance validity testing to ensure the accuracy and reliability of cognitive assessments, particularly for modern digital technologies (Finley, 2024).

In an attempt to address these limitations, it is planned that the Ignite app will undergo development, including translation into multiple languages and the addition of ethnicity data collection. These enhancements will extend the app's reach to a broader, more diverse population, improving the representativeness of the data sets. Additionally, a dot-counting task will be incorporated as a measure of performance validity, which will enhance the app's ability to assess participant effort and ensure the accuracy of cognitive performance results in future assessments. Validation of the Ignite app will be an ongoing process, and we aim to generate normative data that incorporates ethnicity, covers all language versions, and includes the assessment of effort in future studies.

In conclusion, this validation study of Ignite indicates the app can be completed by a broad range of ages and ability levels and provides early evidence that the tests capture cognitive performance reflective of well-established trajectories in normal aging. In addition, the majority of Ignite tests appear to be reliable upon repeated testing, display good concordance with gold-standard neuropsychology tests, and the app is well accepted in healthy controls. One exception is the Time Tap task which will be removed from future versions of Ignite based on the results of this initial study. Future work should continue validation efforts and focus on repeated testing of Ignite through a burst-testing protocol to investigate the extent of practice effects upon multiple administrations. This will be important for clinical trials to define the optimal number of times Ignite should be completed to obtain an accurate depiction of performance. Furthermore, studies should investigate the Ignite app in clinical and preclinical FTD populations to establish if the tests are sensitive to early cognitive impairment. Following further investigation in FTD, the selection of the most sensitive Ignite tests for each genetic group could help to optimize the assessment. It is likely that gene-specific Ignite composite scores would be beneficial to clinical trials in enhancing the sensitivity of detecting cognitive impairment and reducing the required sample sizes. Finally, while Ignite was originally designed for FTD, future work will focus on evaluating its external validity across other neurodegenerative diseases. This will involve assessing its diagnostic accuracy and adaptability to different conditions beyond FTD, ensuring its broader applicability.

References

Adams, T., Pounder, Z., Preston, S., Hanson, A., Gallagher, P., Harmer, C. J., & McAllister-Williams, R. H. (2015). Test–retest reliability and task order effects of emotional cognitive tests in healthy subjects. *Cognition and Emotion*, *30*(7), 1247–1259. https://doi.org/10.1080/02699931.2015.1055713

Barr, W. B. (2003). Neuropsychological testing of high school athletes. Preliminary norms and test–retest indices. *Archives of Clinical Neuropsychology*, *18*(1), 91–101. https://doi.org/10.1016/S0887-6177(01)00185-8

Benussi, A., Alberici, A., Samra, K., Russell, L. L., Greaves, C. V., Bocchetta, M., Ducharme, S., Finger, E., Fumagalli, G., Galimberti, D., Jiskoot, L. C., Le Ber, I., Masellis, M, Nacmias, B., Rowe, J. B., Sanchez-Valle, R., Seelaar, H., Synofzik, M., Rohrer, J. D. … , GENFI Consortium. (2022). Conceptual framework for the definition of preclinical and prodromal frontotemporal dementia. *Alzheimer's & Dementia*, *18*(7), 1408–1423. https://doi.org/10.1002/alz.12485

Binney, R. J., & Ramsey, R. (2020). Social semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience and Biobehavioral Reviews*, *112*, 28–38. https://doi.org/10.1016/j.neubiorev.2020.01.030

Bracken, M. R., Mazur-Mosiewicz, A., & Glazek, K. (2018). Trail making test: Comparison of paper-and-pencil and electronic versions. *Applied Neuropsychology: Adult*, *26*(6), 522–532. https://doi.org/10.1080/23279095.2018.1460371

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. *Psychology and Aging*, *24*(2), 338–348. https://doi.org/10.1037/a0015225

Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex*, *103*, 100–116. https://doi.org/10.1016/j.cortex.2018.02.018

Crawford, J. R., & Garthwaite, P. H. (2008). On the "optimal" size for normative samples in neuropsychology: Capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child Neuropsychology*, *14*(2), 99–117. https://doi.org/10.1080/09297040801894709

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*(1), 45–75. https://doi.org/10.1016/0273-2297(92)90003-K

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really different? *Personality and Individual Differences*, *162*, Article 110045. https://doi.org/10.1016/j.paid.2020.110045

Eslinger, P. J., & Benton, A. L. (1983). Visuoperceptual performances in aging and dementia: Clinical and theoretical implications. *Journal of Clinical Neuropsychology*, *5*(3), 213–220. https://doi.org/10.1080/01688638308401170

Finley, J. A. (2024). Performance validity testing: The need for digital technology and where to go from here. *Frontiers in Psychology*, *15*, Article 1452462. https://doi.org/10.3389/fpsyg.2024.1452462

Ford, K., MacCallum, R., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*(2), 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Franzen, M. D. (2000). *Reliability and validity in neuropsychological assessment*. Springer Science & Business Media. https://books.google.co.uk/books?hl=en&lr=&id=2Nr6-Ss2vh4C&oi=fnd&pg=IA4&ots=roQ9KVX6as&sig=I32u7sCj0Cx3sw_uk6XRV8Ksu9A&redir_esc=y#v=onepage&q&f=false

Goldstein, G., & Watson, J. R. (1989). Test–retest reliability of the Halstead-Reitan battery and the WAIS in a neuropsychiatric population. *Clinical Neuropsychologist*, *3*(3), 265–272. https://doi.org/10.1080/13854048908404088

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Henley, S. M. D., Downey, L. E., Nicholas, J. M., Kinnunen, K. M., Golden, H. L., Buckley, A., Mahoney, C. J., & Crutch, S. J. (2014). Degradation of cognitive timing mechanisms in behavioural variant frontotemporal dementia. *Neuropsychologia*, *65*, 88–101. https://doi.org/10.1016/j.neuropsychologia.2014.10.009

Hoffman, P. (2019). Divergent effects of healthy ageing on semantic knowledge and control: Evidence from novel comparisons with semantically impaired patients. *Journal of Neuropsychology*, *13*(3), 462–484. https://doi.org/10.1111/jnp.12159

Iverson, G. L. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology*, *16*(2), 183–191. https://doi.org/10.1093/arclin/16.2.183

Jiskoot, L. C., Dopper, E. G. P., Heijer, T., Timman, R., van Minkelen, R., van Swieten, J. C., & Papma, J. M. (2016). Presymptomatic cognitive decline in familial frontotemporal dementia: A longitudinal study. *Neurology*, *87*(4), 384–391. https://doi.org/10.1212/WNL.0000000000002895

Jiskoot, L. C., Panman, J. L., van Asseldonk, L., Franzen, S., Meeter, L. H. H., Donker Kaat, L., van der Ende, E. L., Dopper, E. G. P., Timman, R., van Minkelen, R., van Swieten, J. C., van den Berg, E., & Papma, J. M. (2018). Longitudinal cognitive biomarkers predicting symptom onset in presymptomatic frontotemporal dementia. *Journal of Neurology*, *265*(6), 1381–1392. https://doi.org/10.1007/s00415-018-8850-7

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Kemp, J., Després, O., Sellal, F., & Dufour, A. (2012). Theory of mind in normal ageing and neurodegenerative pathologies. *Ageing Research Reviews*, *11*(2), 199–219. https://doi.org/10.1016/j.arr.2011.12.001

Kessels, R. P. C., Montagne, B., Hendriks, A. W., Perrett, D. I., & de Haan, E. H. F. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology*, *8*(1), 75–93. https://doi.org/10.1111/jnp.12009

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. https://doi.org/10.1037/1076-898X.8.2.75

Lemay, S., Bédard, M. A., Rouleau, I., & Tremblay, P. L. G. (2004). Practice effect and test–retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, *18*(2), 284–302. https://doi.org/10.1080/13854040490501718

Lowe, C., Rabbitt, P., & the International Study of Post-Operative Cognitive Dysfunction. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. Cambridge neuropsychological test automated battery. *Neuropsychologia*, *36*(9), 915–923. https://doi.org/10.1016/S0028-3932(98)00036-0

Lynn, R., & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g defined as working memory capacity. *Intelligence*, *36*(3), 226–235. https://doi.org/10.1016/j.intell.2007.06.002

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*(4), 559–571. https://doi.org/10.1017/S0021932000022586

Maylor, E. A., Moulson, J. M., Muncer, A. M., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline in old age? *British Journal of Psychology*, *93*(4), 465–485. https://doi.org/10.1348/000712602761381358

Mayr, U., Spieler, D. H., & Kliegl, R. (2001). *Ageing and executive control*. Psychology Press.

McDowd, J., & Shaw, R. (2000). Attention and aging: A functional perspective. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 221–292). Lawrence Erlbaum.

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: Insights from an fMRI study of the stroop task. *Brain and Cognition*, *49*(3), 277–296. https://doi.org/10.1006/brcg.2001.1501

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moore, K., Convery, R. S., & Rohrer, J. D. (2022). Study protocol: Computerised cognitive testing in a cohort of people with frontotemporal dementia. *BMJ Open*, *12*(8), Article e055211. https://doi.org/10.1136/bmjopen-2021-055211

Moore, K., Convery, R., Bocchetta, M., Neason, M., Cash, D. M., Greaves, C., Russell, L. L., Clarke, M. T. M., Peakman, G., van Swieten, J., Jiskoot, L., Moreno, F., Barandiaran, M., Sanchez-Valle, R., Borroni, B., Laforce, R., Jr., Doré, M. C., Masellis, M., Tartaglia, M. C., … Genetic FTD Initiative, GENFI*. (2022). A modified Camel and Cactus Test detects presymptomatic semantic impairment in genetic frontotemporal dementia within the GENFI cohort. *Applied Neuropsychology Adult*, *29*(1), 112–119. https://doi.org/10.1080/23279095.2020.1716357

Morrison, G. E., Simone, C. M., Ng, N. F., & Hardy, J. L. (2015). Reliability and validity of the neurocognitive performance test, a web-based neuropsychological assessment. *Frontiers in Psychology*, *6*, Article 1652. https://doi.org/10.3389/fpsyg.2015.01652

Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), Article e12217. https://doi.org/10.1002/dad2.12217

Onoda, K., Hamano, T., Nabika, Y., Aoyama, A., Takayoshi, H., Nakagawa, T., Ishihara, M., Mitaki, S., Yamaguchi, T., Oguro, H., Shiwaku, K., & Yamaguchi, S. (2013). Validation of a new mass screening tool for cognitive impairment: Cognitive Assessment for Dementia, iPad version. *Clinical Interventions in Aging*, *8*, 353–360. https://doi.org/10.2147/CIA.S42342

Onoda, K., & Yamaguchi, S. (2014). Revision of the Cognitive Assessment for Dementia, iPad version (CADi2). *PLOS ONE*, *9*(10), Article e109931. https://doi.org/10.1371/journal.pone.0109931

Onyike, C. U., & Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. *International Review of Psychiatry*, *25*(2), 130–137. https://doi.org/10.3109/09540261.2013.776523

Paap, K. R., & Sawi, O. (2016). The role of test–retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002

Paletta, P., Bass, N., Aspesi, D., & Choleris, E. (2022). Sex differences in social cognition. In C. Gibson & L. A. M. Galea (Eds.), *Sex differences in brain function and dysfunction. Current Topics in behavioral neurosciences* (pp. 207–234). Springer. https://doi.org/10.1007/7854_2022_325

Papp, K. V., Rentz, D. M., Maruff, P., Sun, C. K., Raman, R., Donohue, M. C., Schembri, A., Stark, C., Yassa, M. A., Wessels, A. M., Yaari, R., Holdridge, K. C., Aisen, P. S., & Sperling, R. A. (2021). The Computerized Cognitive Composite (C3) in an Alzheimer's disease secondary prevention trial. *The Journal of Prevention of Alzheimer's Disease*, *8*(1), 59–67. https://doi.org/10.14283/jpad.2020.38

Pawlowski, B., Atwal, R., & Dunbar, R. I. M. (2008). Sex differences in everyday risk-taking behavior in humans. *Evolutionary Psychology*, *6*(1), 29–42. https://doi.org/10.1177/147470490800600104

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice*. Prentice Hall.

Possin, K. L., Moskowitz, T., Erlhoff, S. J., Rogers, K. M., Johnson, E. T., Steele, N. Z. R., Higgins, J. J., Stiver, J., Alioto, A. G., Farias, S. T., Miller, B. L., & Rankin, K. P. (2018). The brain health assessment for detecting and diagnosing neurocognitive disorders. *Journal of the American Geriatrics Society*, *66*(1), 150–156. https://doi.org/10.1111/jgs.15208

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Rasmusson, D. X., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the trail making test. *Clinical Neuropsychologist*, *12*(2), 169–178. https://doi.org/10.1076/clin.12.2.169.2005

Rohrer, J. D., Guerreiro, R., Vandrovcova, J., Uphill, J., Reiman, D., Beck, J., Isaacs, A. M., Authier, A., Ferrari, R., Fox, N. C., Mackenzie, I. R. A., Warren, J. D., de Silva, R., Holton, J., Revesz, T., Hardy, J., Mead, S., & Rossor, M. N. (2009). The heritability and genetics of frontotemporal lobar degeneration. *Neurology*, *73*(18), 1451–1456. https://doi.org/10.1212/WNL.0b013e3181bf997a

Rohrer, J. D., Nicholas, J. M., Cash, D. M., van Swieten, J., Dopper, E., Jiskoot, L., van Minkelen, R., Rombouts, S. A., Cardoso, M. J., Clegg, S., Espak, M., Mead, S., Thomas, D. L., De Vita, E., Masellis, M., Black, S. E., Freedman, M., Keren, R., MacIntosh, B. J., … Binetti, G. (2015). Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: A cross-sectional analysis. *Lancet Neurology*, *14*(3), 253–262. https://doi.org/10.1016/S1474-4422(14)70324-2

RStudio Team. (2021). *RStudio: Integrated development for R*. RStudio, PBC.

Ruggeri, K., Maguire, Á., Andrews, J. L., Martin, E., & Menon, S. (2016). Are we there yet? Exploring the impact of translating cognitive tests for dementia using mobile technology in an aging population. *Frontiers in Aging Neuroscience*, *8*, Article 21. https://doi.org/10.3389/fnagi.2016.00021

Salinsky, M. C., Storzbach, D., Dodrill, C. B., & Binder, L. M. (2001). Test–retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12–16-week period. *Journal of the International Neuropsychological Society*, *7*(5), 597–605. https://doi.org/10.1017/S1355617701755075

Sanders, L. M. J., Hortobágyi, T., Balasingham, M., Van der Zee, E. A., van Heuvelen, M. J. G., Maria, L., & Sanders, J. (2018). Psychometric properties of a flanker task in a sample of patients with dementia: A pilot study. *Dementia and Geriatric Cognitive Disorders Extra*, *8*(3), 382–392. https://doi.org/10.1159/000493750

Seeley, W. W., Crawford, R., Rascovsky, K., Kramer, J. H., Weiner, M., Miller, B. L., & Gorno-Tempini, M. L. (2008). Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. *Archives of Neurology*, *65*(2), 249–255. https://doi.org/10.1001/archneurol.2007.38

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Singh, V., Thakral, S., Singh, K., & Garg, R. (2022). Examining cognitive sex differences in elite math intensive education: Preliminary evidence from a gender inequitable country. *Trends in Neuroscience and Education*, *26*, Article 100172. https://doi.org/10.1016/j.tine.2022.100172

Staffaroni, A. M., Quintana, M., Wendelberger, B., Heuer, H. W., Russell, L. L., Cobigo, Y., Wolf, A., Matt Goh, S.-Y., Petrucelli, L., Gendron, T. F., Heller, C., Clark, A. L., Carson Taylor, J., Wise, A., Ong, E., Forsberg, L., Brushaber, D., Rojas, J. C., VandeVrede, L., … Afonso, S. (2022). Temporal order of clinical and biomarker changes in familial frontotemporal dementia. *Nature Medicine*, *28*(10), 2194–2206. https://doi.org/10.1038/s41591-022-01942-9

Tombaugh, T. N. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Tsai, R. M., & Boxer, A. L. (2016). Therapy and clinical trials in frontotemporal dementia: Past, present, and future. *Journal of Neurochemistry*, *138*(Suppl. 1), 211–221. https://doi.org/10.1111/jnc.13640

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Suppl. 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

White, N., Forsyth, B., Lee, A., & Machado, L. (2018). Repeated computerized cognitive testing: Performance shifts and test–retest reliability in healthy young adults. *Psychological Assessment*, *30*(4), 539–549. https://doi.org/10.1037/pas0000503

Wu, W., & Hoffman, P. (2022). Validated measures of semantic knowledge and semantic control: Normative data from young and older adults for more than 300 semantic judgements. *Royal Society Open Science*, *9*(2), Article 211056. https://doi.org/10.1098/rsos.211056

Received June 17, 2024
Revision received January 9, 2025
Accepted January 9, 2025 ■