Contemporary Clinical Trials 78 (2019) 76–87

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Intervention Comparative Effectiveness for Adult Cognitive Training (ICE-ACT) Trial: Rationale, design, and baseline characteristics

Jong-Sung Yoon[a,*], Nelson A. Roque[b], Ronald Andringa[a], Erin R. Harrell[a], Katharine G. Lewis[a], Thomas Vitale[a], Neil Charness[a], Walter R. Boot[a]

[a] Florida State University, Department of Psychology, Tallahassee, FL, United States
[b] Penn State University, Center for Healthy Aging, University Park, PA, United States

ARTICLE INFO

Keywords:
Clinical trial
Comparative effectiveness
IADL
Cognitive intervention

ABSTRACT

Age-related perceptual and cognitive declines are associated with difficulties performing everyday tasks required to remain independent. Encouraging improvements in cognitive abilities have been shown for various short-term interventions but there is little evidence for direct impact on independence. This project compares the effect of broad and directed (narrow) technology-based training on basic perceptual and cognitive abilities in older adults and on the performance of simulated tasks of daily living including driving and fraud avoidance. Participants (N = 230, Mean age = 72) were randomly assigned to one of four training conditions: broad training using either (1) a web-based brain game suite, Brain HQ, or (2) a strategy video game, Rise of Nations, or to directed training for (3) Instrumental Activities of Daily Living (IADL) training using web-based programs for both driving and fraud avoidance training, or (4) to an active control condition of puzzle solving. Training took approximately 15–20 h for each intervention condition across four weeks. Before training began, participants received baseline ability tests of perception, attention, memory, cognition, and IADL, including a driving simulator test for hazard perception, and a financial fraud recognition test. They were tested again on these measures following training completion (post-test). A one-year follow-up from training completion is also scheduled. The baseline results support that randomization was successful across the intervention conditions. We discuss challenges and potential solutions for using technology-based interventions with older adults. We also discuss how the current trial addressed methodological limitations of previous intervention studies.
Trial registration number: NCT03141281

1. Introduction

Age-related perceptual and cognitive declines are associated with difficulties performing everyday tasks required to remain independent [1]. Encouraging improvements in cognitive abilities have been shown for a variety of short-term interventions such as group and individual training on abilities (e.g., the ACTIVE trial; [46]), digital games [6,23], aerobic exercise [26], and combined treatments [5]. The rationale for these interventions is that they can be expected to support the maintenance of functional independence in older adulthood given the relatively strong relationship between Instrumental Activities of Daily Living (IADLs) performance and cognitive abilities [1], and the range of IADLs required for independence.

There are, however, several limitations to the design and interpretation of broad ability-level cognitive training described above [43]. First, over a century of research on skill acquisition has shown relatively narrow transfer and strong domain specificity for acquired skills [20]. In fact, sometimes a narrow training package may be more effective than training of general abilities, namely broad training, depending on the type of task and target outcome [33]. Given the range of interventions that have shown some efficacy, application of this knowledge to deploying cost-effective population-level interventions requires assessing their comparative effectiveness. Second, relatively few intervention studies have assessed independence using metrics other than cognitive ability measure proxies. For example, many studies have used laboratory task outcomes such as the Useful Field of View (UFOV) test - as a predictor of driving crashes [22] - rather than using performance data from real or simulated driving. Third, practice effects on cognitive measures that are repeated from baseline to post-training test can make it difficult to estimate true change, although increased participant age tends to minimize practice effects and alternate form use can almost eliminate practice effects for some tests [11], though not

* Corresponding author.
E-mail address: yoon@psy.fsu.edu (J.-S. Yoon).

https://doi.org/10.1016/j.cct.2019.01.014
Received 1 November 2018; Received in revised form 20 January 2019; Accepted 30 January 2019
Available online 31 January 2019
1551-7144/

for all tasks during relatively long periods of intervention [e.g., [37–39]]. Finally, placebo effects, particularly expectation effects, are a serious threat to interpreting intervention studies [8].

This randomized trial was designed to compare the effectiveness of broad and directed (narrow) technology-based training on basic perceptual and cognitive abilities in older adults and on the performance of simulated tasks of daily living. Participants were randomly assigned to four training conditions: (1) a brain game; (2) a video game; (3) IADL training for both fraud avoidance and driving; or (4) an active control condition, involving puzzle solving. This trial seeks to provide guidance not just about what works, but also what works best for a given investment (i.e., time, money) in older adults. The current trial was specifically designed to address the several methodological limitations mentioned above, including: (1) the development of direct measures of IADL performance, targeting areas related to risks for catastrophic failure (e.g., crashes in driving, losing financial resources to fraud); (2) the use of alternate forms to provide more precise estimates of treatment effects (for both cognitive and IADL measures); (3) inclusion of an active control condition; and (4) measurement of expectation effects (as recommended in [43]). In this paper, we provide a primer to the design of this trial and describe the baseline characteristics of our sample.

2. Methods

2.1. Overview of the design

This trial is registered on ClinicalTrials.gov (Identifier#: NCT03141281). The trial is a single-site randomized controlled trial (RCT), conducted at Florida State University (FSU; Tallahassee, FL), one of four funded sites of the National Institute of Aging Center for Research and Education on Aging and Technology Enhancement (CREATE). Following a telephone screening and baseline assessments, eligible participants were randomized into one of four conditions: (1) Posit Science's BrainHQ; (2) Rise of Nations, a PC-based video game; (3) Instrumental Activities of Daily Living (IADL) training, on both AARP's Driver Safety Course, and a Financial Fraud training course developed for this intervention; (4) active control group, playing a suite of three tasks (i.e., Sudoku, crossword puzzles, and word search). Once participants were assigned to a condition, they received a laptop with specialized software for their condition. Participants received in-lab training designed to encourage familiarity with how to operate a laptop, and the necessary components for their task. Participants were asked to train on their assigned condition for four weeks at home, and return for two scheduled days of post-training assessments, as well as a follow-up assessment at 12 months. The overall study design and contact schedule is summarized in Fig. 1. The trial had highly standardized protocols for recruitment, screening, assessment, intervention administration (including manuals sent home with participants), and data transfer from intervention devices. The Institutional Review Board (IRB) at Florida State University, approved the protocol for this trial.

2.2. Eligibility criteria

Participants were considered eligible for the study if they: (1) were 65 years of age or older; (2) planned to stay in the Tallahassee, Leon County Area (Florida, USA) for the next 12 months; (3) had a valid driver's license and drove at least once a month; (4) had adequate cognitive ability, assessed via a phone interview using the Wechsler Memory Scale III (story A score of ≥ 6 or story B score of ≥ 4 if they failed story A). Participants were also required to speak English.

Participants were ineligible if they: (1) reported terminal illness with life expectancy < 12 months; (2) reported or exhibited a disabling visual condition assessed as the inability to read printed material; (3) reported or exhibited a disabling speech hearing and comprehension condition assessed by inability to hear and comprehend the screener's instructions; (4) reported or exhibited a disabling speech production condition assessed as the inability to respond with comprehensible English speech to the screener's queries; (5) reported or exhibited a disabling psychomotor condition assessed as the inability to use a keyboard and pointing device.

Participants were also ineligible if they had ever: (1) completed the AARP Driver Safety course (https://www.aarp.org/auto/driver-safety); (2) played the Rise of Nations video game (https://en.wikipedia.org/wiki/Rise_of_Nations); (3) trained with Posit Science's Brain HQ platform (https://www.brainhq.com); or (4) trained with Aptima's Mind Frontiers video game, as part of a previous intervention [44]. We also assessed previous experience with such games using the previous experience questionnaire described in Table 1.

2.3. Intervention conditions

2.3.1. The BrainHQ condition (BHQ condition)

2.3.1.1. Overview. Participants in the BHQ condition completed a subset of the brain training exercises available on the BrainHQ platform developed by Posit Science (i.e., 3 of the 22 tasks available at the time of writing; https://www.brainhq.com). Training in this study focused on speed of processing, and consisted of practicing computer-based exercises that involved detecting, identifying, discriminating, and localizing aspects of often briefly presented targets [e.g., [4]]. The three tasks completed by participants were: Double Decision, Freeze Frame, and Target Tracker (described in more detail below). The tasks in BrainHQ are gamified exercises analogous to laboratory assessment and training tasks, with Double Decision analogous to UFOV, Freeze Frame analogous to a choice reaction time task, and Target Tracker analogous to the dynamic multiple object tracking task [(e.g., [31]). These tasks were chosen to explore the nature of near and far transfer for speed of processing training. Previous studies have suggested that speed of processing training might demonstrate far transfer to improved performance on timed IADL and some driving measures [3,17,35].

2.3.1.2. Double decision. In this task, a vehicle was briefly, centrally presented, followed by the brief presentation of a 'Route 66' sign in the periphery of the screen. Participants were instructed to select which vehicle was displayed from two options, and subsequently report the location of the 'Route 66' sign among eight radial wedges.

2.3.1.3. Freeze frame. Participants were shown a target image to memorize before each level of trials. After memorization of the target image, participants were shown images sequentially. If the presented image was not the target image, participants responded by pressing the 'right arrow' key. If the presented image was the target image, participants were instructed to withhold a keyboard response (i.e., "Freeze").

2.3.1.4. Target Tracker. Participants were instructed to track a set of target items as they moved randomly across the screen, while ignoring similarly moving distractor items. Once the objects stopped moving, participants were instructed to report the final location of all target items by clicking on each (now-masked) object.

2.3.1.5. In-lab training session. During in-lab training, participants assigned to this intervention condition were trained on how to complete BHQ software training. Participants completed the in-lab training session tutorial guided by study personnel. Participants were trained on: (1) how to access the BHQ software; (2) how to interact with the BHQ software to complete their assigned tasks; and (3) specific task mechanics. Participants received a guided explanation of the training schedule as well. Training on each BHQ task lasted about 10 min for a total of 30 min per training session. Each task used an adaptive staircase to adjust the difficulty level of the task, across tasks, based on the participants' ability.

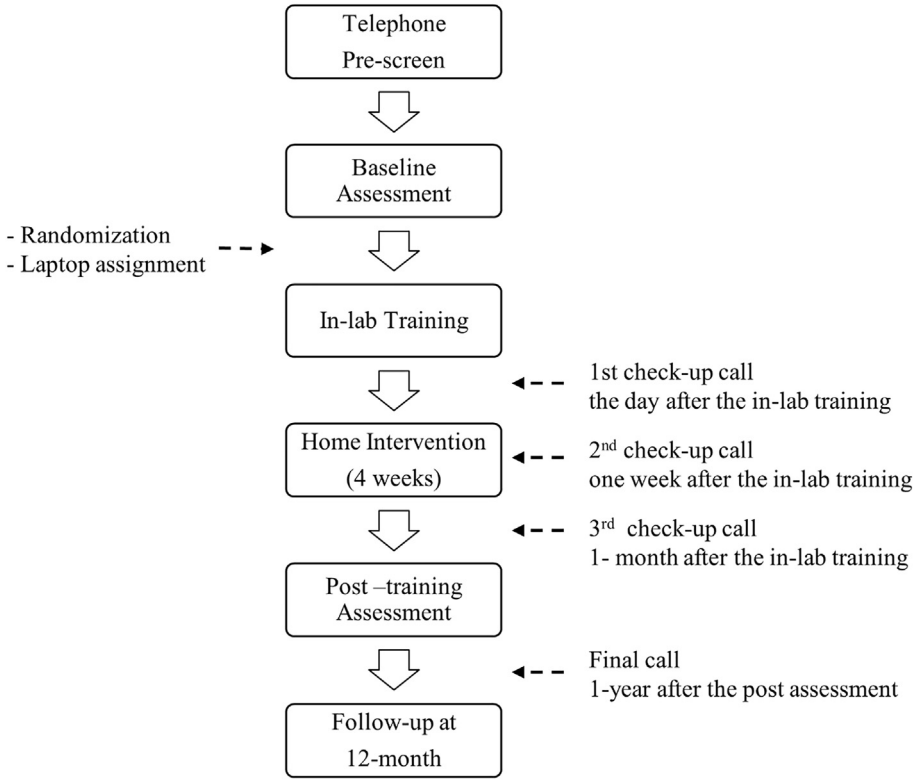

Fig. 1. Overall study design and contact schedule.

2.3.1.6. Home-based intervention. After in-lab training participants were instructed to complete two training sessions a day, for five days a week, for a total of four weeks (20h) in their home. Participants were instructed to play a minimum of 30 min per training session, for a minimum total of one hour each day. Each session followed the previous training session's adaptive difficulty.

2.3.2. The Rise of Nations (RON) condition

2.3.2.1. Overview. Rise of Nations®: Gold Edition, developed by Big Huge Games and published by Microsoft Game Studios, is a real-time strategy game in which players are tasked with leading a nation from the ancient age to the information age. Players start with a group of citizens controlling a small portion of a larger map. Throughout gameplay, players discover and acquire resources to build more and larger cities, expand their territory, create an army for offense and defense, develop new technologies, and progress their nation through historical ages. Other computer-controlled players attempt to accomplish these same goals on the map, resulting in competition for territory and resources. There are multiple ways for players to win the game: territory victory, military victory, and wonder victory. Territory victories are achieved when players gain control of 70% of the map by expanding their national borders. Players achieve military victories by capturing an opponent's capital city. Wonder victories are a nonviolent victory condition achieved by constructing unique and noteworthy structures.

2.3.2.2. Game settings. Participants were set to play “Quick Battle” games, a single player game mode with customizable settings. Participants competed against a single computer opponent whose difficulty level was set to “easiest.” Participants and computer-controlled players' nationalities were randomized. Participants' teams and allies were set to “Survival of the fittest”, namely, a mode in which no collaboration is allowed between nations. Game play Participants were asked to play at “Arena” - a small map spotted with large lakes – where allows participants to use land, naval, and air units to achieve a victory. Participants were positioned closer to their opponents throughout gameplay to encourage the completion of a victory condition and reduce the length of time needed to find opponents.

2.3.2.3. In-lab training session. During in-lab training, participants assigned to this intervention condition were trained on how to complete RON training. Participants completed the first game training session tutorial guided by study personnel. Participants were trained on how to access the Rise of Nations software and manipulate game settings for their training session. Participants were provided with an overview of how to control and interact with the RON software. Participants were guided with explanations of various gameplay mechanics. Participants acquired resources, built multiple cities, expanded their territory, created an army for offense and defense, developed new technologies, and progressed their nation through two historical ages. Participants received guidance on how each game mechanic is applied to gameplay strategies to complete each of the three victory conditions: territory victory, military victory, and wonder victory. Participants were also instructed on how to: (1) complete a training session; (2) save their current game; and (3) load a saved game to continue a training session.

2.3.2.4. Home-based intervention. After in-lab training participants were instructed to complete two training sessions a day, for five days a week for a total of four weeks (20 h) in their home. Participants were instructed to play a minimum of 30 min per training session, for a minimum total of 1 h each day. Participants started their training on the easiest difficulty level. At the end of each session, participants saved game progress to their profile, and participants were instructed to begin the next training session at the previous training session's save point. Participants were instructed to try to play a game of RON during each training session until they achieved victory or defeat. Participants were instructed to manually adjust the difficulty level of the computer

Table 1
Description of the assessment batteries.1

Measure		Name of measure	Description
Screening measures		Telephone Pre-screening	Participants were asked to answer basic demographics (age, gender, & race) along with inquiries about eligibility criteria. [R]
		Wechsler Memory Scale III [49]	This was given during the phone screening to judge adequate cognitive ability. Participants were asked to recall elements from brief stories. [R]
Demographic information		Demographic Questionnaire	Included more questions about demographic characteristics (e.g. education, income, housing, occupation). [B]
Primary outcome	IADL	Fraud Detection Task	Participants received text-based vignettes and were asked to rate how confident they were the vignette was an attempt of fraud using 0–100 scale (100% confident that the scenarios are a fraud attempt). Each set consist of six non-fraud and four fraud attempts and three parallel sets were generated. [B, P, F].
		Driving Simulator Hazard Perception Task	Participants were asked to drive scenarios in the DriveSafety simulator. Each scenario exposes participant to hazardous situations that allow us to evaluate their driving skills with various measures, such as velocity, lane position (deviation from the middle of lane: −1 [left] ~ 1 [right]), max brake force (0–1 point), and number of collisions. We generated three parallel scenarios. [B, P, F]
		Self-reported IADL Difficulty	We used a short-form from Multicenter AIDS Cohort Study (MACS), which is a modified version of the Lawton IADL [30]. Items asking difficulties in completing 14 different instrumental activities of daily living. Seven of those activities (housekeeping, managing finances, telephone use, cooking, transportation, shopping, taking medication) were selected and scored on 0–3 scales (0: no difficulty – 3: severe difficulty) similar to ACTIVE trial [50] (Cronbach's $\alpha = 0.83$). [B, P, F]
	Speed of processing	Useful Field of View [UFOV; [16]]	A computer-administered test of functional vision and visual attention. The test consists of three subtests measuring processing speed, divided attention, and selective attention. A summed score of the subsets was reported. [B, P, F]
		Digit Symbol Substitution Test [DSST; [48]]	A pencil and paper based test of processing speed. Participants received 9 digit-symbol pairs followed by lists of random digits. During 90 s, participants were asked to write down the corresponding symbol under each digit as fast as possible. [B, P, F].
	IADL training knowledge	IADL Training Knowledge Assessment -Driving	Three multiple choice questions asking about specific driving knowledge presented in the AARP Driver Safety Course. We generated two parallel sets. [P, F]
		IADL Training Knowledge Assessment - Finances and Fraud	Three multiple choice questions asking about specific knowledge about a type of fraud and how to manage finances presented in the Finances and Fraud education. We generated two parallel sets. [P, F]
Secondary outcome	Technology proficiency	Computer Proficiency Questionnaire [CPQ; [7]]	12 items that assess proficiency of 6 different computer related skills (basics, printing, communication, internet, scheduling, & multimedia) on 5-point scale (1 = never tried – 5 = very easily) (Cronbach's $\alpha = 0.87$). For each skill, answers for all items were averaged. The averaged scores were then summed across 6 different skills to come up with a total proficiency measure. [B, P, F]
		Mobile Device Proficiency Questionnaire [MDPQ; [36]]	16 items that assess proficiency of 8 different mobile device related skills (basics, communication, storage, internet, calendar, entertainment, privacy, & troubleshooting and software management) on 5-point scale (1 = never tried – 5 = very easily) (Cronbach's $\alpha = 0.95$). Like CPQ, for each skill, answers for all items were averaged. The averaged scores were then summed across 8 different skills to come up with a total proficiency measure. [B, P, F]
	Numeracy	Numeracy Test	7 items that assess numeracy. We used three parallel forms selected from [12,41]. [B, P, F]
	Reasoning	Raven's Advanced Progressive Matrices [RAPM; [34]]	The RAPM present participants with a complex visual pattern with a piece cut out of it. Participants were asked to find the missing piece that completed the pattern. The full version of the RAPM (36 items) was divided into three parallel sets of approximately equal difficulty based on [2]. Participants received 20 min to complete each set (12 items) at each assessment. [B, P, F]
		Letter sets [18]	Participants received five of the letter sets and were asked to find the rule which related four of the letter sets to each other and to mark the one which does not fit the rule. Participants received 10 min to complete 15 items. We used three parallel sets. [B, P, F].
	Memory	Hopkins Verbal Learning Test [HVLT; [10]]	The HVLT consists of three trials of free-recall of a 12-item, semantically categorized list, followed by a delayed recall trial (20–25 min delay). During the three trials, participants were instructed to listen carefully as the examiner reads the word list and then were asked to recall as many words as possible. We used three parallel sets. [B, P, F].
		Rey Auditory Verbal Learning Test [RAVLT; [40]]	The RAVLT presented a list of 15 words across five trials and at the end of each presentation participants were asked to recall as many words as possible. After the five trials, a second list of 15 words was presented to the participants and they were asked to recall the words from the second list (trial 6). Immediately following trial 6, the participants were again asked to recall the words from the first list (trial 7). After a 20-min delay, the participants were asked to recall the words from the first list. We used three parallel sets. [B, P, F]

(*continued on next page*)

Table 1 (*continued*)

Measure		Name of measure	Description
	IADL	University of Miami Computer-Based Functional Assessment Battery (UMCFAB; [14])	The UMCFAB was designed to assess performance on a variety of everyday activities using computer-based simulation. We used two tasks: ATM banking and prescriptions refill. The ATM task was a replication of a current operational ATM system and the prescription refill task was a simulation of a local chain pharmacy. The ATM task asked participants to check the balance in their savings account, transfer money between checking and savings accounts, and to withdraw cash from a savings account. The prescription task asked participants to call the pharmacy (using a telephone keypad on the screen), refill two different prescriptions, and request a pickup time and date. A rate measure (total correct responses/time) - reflecting task efficiency - was used as a main index. [B, P, F]
	Perception of training effectiveness	Post-training Questionnaire	The questionnaire consists of seven items asking about participants' perception of training effectiveness on motor, cognitive, and IADL abilities based on 7-point scale (e.g., "Training activities like the one I was given to complete have the potential to improve reasoning ability": 1 = very strongly disagree – 7 = very strongly agree). [P]
	Previous Experience	Previous Experience Questionnaire	Included more questions about participants' possible experiences of fraud, IADL training, cognitive training, and video games in the last 12-month. [B, F]

[1] R: During recruitment, B: Baseline, P: Immediate post-assessment, F: 12-month follow-up.

opponent after completion of a game. If participants achieved two victories in a row they were instructed to increase the difficulty by one level. If participants lost, they were instructed to reduce the difficulty by one level. Participants were also instructed to read the take home manual if they required tips on how to complete the home training session, adjusting difficulty settings after completing a game, or on how to use training hardware.

2.3.3. IADL training

2.3.3.1. Overview. Participants assigned to this intervention condition were asked to complete two different online IADL tasks: 1) AARP Driver Safety Course (https://www.aarpdriversafety.org) and 2) Finances and Fraud education. The order of IADL tasks was counterbalanced across participants in this group.

2.3.3.2. On-line courses. The AARP driver safety course was designed to refresh an older adults' driving skills and knowledge of safe driving strategies. The course consisted of six units, each ending with a short quiz to assess participants' knowledge of the module. At the end of the course (i.e. the last unit), participants completed a review module and a post-course quiz. Participants moved linearly through the course, clicking a 'Next' button to advance to the subsequent section, with the option to revisit slides already completed.

The 'Finances and Fraud' education was designed to help participants learn how to manage finances and protect themselves from various types of financial and online fraud. The user interface and level of interactivity of the 'Finances and Fraud' education course closely mimicked the interface of the AARP course, through the use of a common e-learning platform (i.e., Articulate Storyline2; https://articulate.com/). The course consisted of eight units, each ending with a short quiz to assess participants' knowledge of the module. At the end of the course (i.e. the last unit), participants completed a review module and a post-course quiz. Participants moved linearly through the course, clicking a 'Next' button to advance to the subsequent section, with the option to revisit slides already completed.

2.3.3.3. In-lab training session. During in-lab training, participants received individualized training on how to complete both online courses using their assigned laptop. During in-lab training, participants were instructed to complete the first unit of either IADL course (i.e., 'Finances and Fraud', AARP Driver Safety; depending on their counterbalanced condition) at their own pace.

2.3.3.4. Home-based intervention. After the in-lab training, participants were asked to complete each course one at a time by spending around 30–60 min per day, for five days a week. Once they completed both courses for the first time, given that each course took around three to 5 h to complete, participants were asked to repeat both courses to enhance learning and retention. As a result, the total number of training hours were approximately equivalent to the other intervention conditions (i.e., participants completed each course twice within the next four weeks after baseline assessment. Note that participants were asked not to complete the last unit of each course during the 1st cycle given that the last unit was a review module with a post-course quiz.

2.3.4. The Active control condition

2.3.4.1. Overview. The active control condition was a software suite consisting of three computerized puzzle games. The active control condition presented participants with three different training tasks to complete per training session: Desktop Crossword game (developed and published by Inertia software), Britannica Sudoku Unlimited game (developed and published by Britannica Games), and Britannica Word Search game (developed and published by Britannica Games).

2.3.4.2. Crossword. Participants completed a computerized version of traditional paper-based crossword puzzles. Participants filled in blank squares on the playing board with specific words, using the list of clues. Participants then used the letters that are in place to fill out the spaces based on their previous answers, with the constraint that overlapping words on the board must share letters.

2.3.4.3. Sudoku. Participants completed a computerized version of the traditional paper Sudoku game. Participants filled in blank squares on grid that is usually 9 × 9. The grid is further divided into smaller grids called boxes that are 3 × 3. A subset of squares begin with numbers between one and nine filled in. Participants solved this puzzle by filling in each row and column with a number between one and nine into the grid, with the constraint that each number could only appear in each row, column, and box once.

2.3.4.4. Word search. Participants completed a computerized version of the traditional paper word search game. Participants, given a list of words, must figure out where the corresponding words are in a matrix comprised of a jumbled selection of letters. The words could be placed vertically, horizontally, diagonally, or backwards. Participants clicked and dragged with a mouse to highlight and remove a word from the

given list.

2.3.4.5. In-lab training session. During in-lab training, participants assigned to this intervention condition were trained on how to complete active condition software training. Participants completed the in-lab training session tutorial guided by study personnel. Participants were trained on: (1) how to access the three software packages comprising the Control condition; (2) how to interact with the active control condition software; and (3) various task mechanics and rules. Participants trained on word search for 15 min, crossword for 15 min, and Sudoku for 30 min. Participants were given a longer period of in-lab training on Sudoku due to the complex rule set of Sudoku.

2.3.4.6. Home-based intervention. After in-lab training participants were instructed to complete two training sessions a day, for five days a week for a total of four weeks (20 h) in their home. Unlike the in-lab training, participants were instructed to play a minimum of 30 min per training session, for a minimum total of 1 h each day. Across the entire at home portion of the study participants should train on Sudoku a total of 7 h, word search a total of 6.5 h, and crosswords a total of 6.5 h.

2.4. Procedures

2.4.1. Pre-screening

Participants were contacted over the phone and given a brief description of the requirements and design of the project. If participants were 65 or older and interested in participation, we would proceed with the 10-minute pre-screening process. Participants were asked basic demographic questions along with inquiries about any potential deficits with their vision, hearing, and physical mobility. Participants were read Story A from the Wechsler Memory Scale II. If participants could not repeat six or more elements from the story then they were read Story B. If participants still could not repeat four or more elements from Story B, then they did not pass the initial criteria. If participants did pass the initial prescreen, we asked specific questions to determine if they were eligible for the study. Participants were asked if they planned on moving away from the Leon County, FL, USA area within the next year or if they had a terminal illness with the life expectancy less than one year. If they answered yes, this would make them ineligible for the study, which requires a follow up appointment after one year. Participants were also asked if they had a valid driver's license and if they drove at least once a month. We then asked participants if they had played or completed any of the following programs: AARP Driver Safety Course, Rise of Nations, BrainHQ, or Aptima's Mind Frontiers video game (used in a previous intervention at FSU). Completing any of these programs would make them ineligible for the study. All eligible participants were contacted within a week of the pre-screening to schedule lab visits for baseline assessment and laptop training.

2.4.2. Contact schedule

Participants were scheduled to be contacted four times throughout the Comparative Effectiveness study. The first call was the day after their in-lab training. This call was to: (1) make sure they had no issues connecting to their home WiFi (if applicable); (2) remind participants about the nature of their assigned condition; and (3) to answer any questions the participant might have about the laptop tasks. The second call was one week after the participant had received their laptop. We asked participants if they had any hardware or software issues with the laptop. The third call was made one month after laptop training and the purpose was to confirm the days and times of their two post-testing appointments in the lab. The final call is scheduled to be made one year after the participants' post testing appointments, in order to schedule their two follow-up appointments to the lab.

Each participant initially came into the lab for three days, this included two days of baseline assessments and one day of laptop training. Each appointment lasted approximately 2 h, with a five-minute break during both baseline testing days. Participants were paid $20 at the end of each appointment.

2.4.3. Baseline assessment - day 1

After Participants signed and dated the consent form they were asked to complete a background demographic survey and fraud survey. In order to ensure randomization of task versions (parallel forms), a randomization code was placed on each participant's folder (e.g., ABC, CBA, BAC, etc.), that corresponded to the task version participants were to complete at each assessment point (e.g., if code was ABC, A = form for all tasks at baseline, B = form for all tasks at post-testing, C = form for all tasks at follow-up). This code was typed into an online survey that we generated based on Qualtrics (https://www.qualtrics.com/), and using display logic, participants were presented with the version of the assessments corresponding to their randomization condition (for more detail, see Section 2.7.3. Stepwise protocol in Qualtrics).

Participants were then given the Hopkins Verbal Learning Test (HVLT; [10]). Participants listened to a recording of 12 words in three semantic categories and were asked to repeat back as many as they could remember. Trained research assistants circled the words the participant repeated back and included any intrusions the participant might have said. This recording was played three times, and then a fourth time as part of the delayed recall task approximately 20 min later. Participants were then given 10 min to complete the Letter sets task [18]. Each problem had five sets of letters with four letters in each set and the participants had to determine which four of the sets were alike, and select the one set that was different from the rest. Participants then completed three questionnaires on the computer: the Computer Proficiency Questionnaire (CPQ-12; [7]), the Mobile Device Proficiency Questionnaire (MDPQ; [36]), and the Wireless Network Proficiency Questionnaire (WNPQ) and Wireless Network Demographics Questionnaire (in preparation for publication). Following the questionnaires, participants were given the Berlin Numeracy task [12,41]. When this was completed the Hopkins Delayed Recall task was administered, participants were asked to repeat any words they remembered from the list. Immediately following was the Delayed Recognition (Forced Choice) trial. Participants were then give a five-minute break. After the five-minute break participants completed the Fraud Scenario Task and the Instrumental Activities of Daily Living (IADL) Short Form Task. After these surveys the participants completed the E-Prime computer program designed to administer Raven's Advanced Progressive Matrices (RAPM; [34]). After the RAPM we reminded participants of their Baseline Day 2 appointment, paid them $20, and thanked them for their time.

2.4.4. Baseline assessment - day 2

The first task for Baseline Day 2 was the Rey Auditory Verbal Learning Task (RAVLT; [40]). Participants were played a voice recording of 15 words for five trials and asked to repeat what they remembered. Research assistants circled the words the participant repeated back and included any intrusions the participant might have said. A voice recording then read off 15 different words for Trial B that they were asked to repeat back as many as they could remember. Then participants were asked to repeat any of the words they remembered from the first list. This task did include a Delayed Recall portion that was given to participants at a later time during the appointment. Participants were then given the Digit Symbol Substitution Task (DSST; [48]) which required them to complete as many symbols in 90 s as they could. Participants were then placed at a computer and asked to complete the UFOV task which measured their visual attention, followed by the delayed recall portion of the Rey Auditory Verbal Learning Task in which participants were asked to repeat back any words they remember from the list. Participants were then set up with another laptop only specialized to administer a simulated IADL task, namely, University of Miami Computer-Based Functional Assessment Battery (UMCFAB; [14]). In the UMCFAB, participants were specifically asked to complete

simulated banking and prescription refill tasks.

After these tasks participants were led to a separate room where a driving simulator was housed. Before participants completed any driving simulator task they were asked the following questions regarding simulator sickness: 1) Would you rate any discomfort in your head, including your eyes, as none, slight, moderate, or severe?; 2) How about stomach discomfort-would you rate that as none, slight, moderate, or severe?; 3) And any discomfort any place else- would you rate that as none, slight, moderate, or severe? Research assistants were instructed to end the task immediately if participants reported they were in too much discomfort to continue the task. Participants were also told to alert the research assistants at any time if they felt discomfort (dizziness, headache, nausea, etc.) and allowed to stop or opt out any following simulator tasks due to the simulator sickness. Participants then started the Left-Hand Judgement task in which they were stationed at a green light and had to indicate by pressing the red steering wheel button each time they felt it was safe to proceed with a left-hand turn. Following the task research assistants asked participants the three simulator sickness questions. The research assistants then had participants go through the Driving Familiarization Task so they could practice working with the simulator controls and again asked them the three simulator sickness questions. The final task was the Hazard Perception Task which had participants go through a full driving scenario on their own guided by a GPS system. The research assistants then asked participants the three simulator questions again, reminded them of their In-lab Training appointment and compensated them $20.

2.4.5. In-lab training

The participants began their in-lab training day with a brief laptop hardware training session. During this session a research assistant set up the participant's assigned laptop with their randomized condition. Once participants were assigned to a condition, a laptop was configured with software appropriate to their assigned condition, through a suite of custom created Python and Batch file scripts. The laptop assignment process also prepared data logging scripts to track: (1) logon and logoff times; (2) shutdown and startup times; (3) idle time — a screensaver timeout was set to one minute; (4) usage of the assigned condition; (5) usage of the session diary. In addition to these measures, at the end of each daily session (i.e., when participants closed the program for a task), a session diary consisting of six questions was set to automatically open, so that participants could provide their feedback. The first four questions were statements about the daily session in which participants had to choose an option on a 7-point Likert-scale (1 = strongly disagree; 7 = strongly agree). The four questions were: (1) I found today's session to be enjoyable; (2) I found today's session to be challenging; (3) I found today's session to be frustrating; (4) I was motivated to perform well on today's session. The last two questions were: (1) during today's session, for how many minutes, did you stop to complete another task?; (2) do you have any comments about today's session?

After participants were shown the basic hardware features of the laptops, they were trained on their specific condition as outlined in the sections above (i.e., Section 2.2, In-lab training session). A training script was created for each of the conditions. The participants were trained on their specific tasks from start to finish so that they were familiar with the order of activities for the next month of at-home laptop training. This initial process took approximately 1 h. Next, for 1 h, participants were asked to complete the training on their own with the research assistant in the room in case they had any questions. At the end of the appointment, participants were given a manual to take home, which included the training schedule, the lab's contact information, step-by-step instructions for their assigned condition, and a hardware manual. Upon their departure after this assessment, participants were given: (1) a laptop with a charger; (2) a bag for the laptop; (3) a USB laser mouse; and (4) a 20 ft. Ethernet cord. Participants who did not have home internet were provided a MiFi device with 2 GB of data preloaded on it.

2.4.6. Home-based intervention

Participants completed the tasks of their assigned condition as outlined in the sections above (i.e., Section 2.2, Home-based intervention).

2.4.7. Post-training assessment - day 1

Participants were scheduled to return to the lab for post-assessment testing four weeks from their laptop training day. Participants returned the laptop and all other materials to the lab when they arrived. Procedures of Post-assessment Day 1 followed the same order and tasks as those of Baseline Day 1 except for two differences: (1) participants did not have to complete the background questionnaire or the Fraud Survey; (2) participants completed the IADL Training Knowledge Assessment following the five-minute break.

2.4.8. Post-training assessment - day 2

Procedures of Post-assessment Day 2 followed the same order and tasks as those of Baseline Day 2. The only difference depended on how the participant felt during the driving simulator tasks they previously completed on Baseline Day 2. If the participant had experienced simulator sickness during the baseline assessment, we would not require them to do the Driving Familiarization Task or the Hazard Perception Task for Post-assessment Day 2. The participants would still be allowed to complete the Left-Hand Judgement task because the task was stationary.

2.4.9. 12-month follow-up - day 1

Procedures of the 12-month Follow-up Day 1 were scheduled to follow the same order and tasks as those of Baseline Day 1 except for two differences: (1) the participants did not complete the demographic questionnaire; (2) participants completed the Near Transfer Task that was administered during their Post-training Assessment Day 1.

2.4.10. 12-month follow-up - day 2

Procedures of the 12-month Follow-up Day 2 were also scheduled to follow the same order and tasks as those of Baseline Day 1 and Post-training Assessment Day 2.

2.5. Measures

All outcome measures are further outlined in the sections below. For detailed explanations of each outcome measure, see Table 1. The measures and analysis plans are also detailed on ClinicalTrials.gov (https://clinicaltrials.gov/ct2/show/NCT03141281) and on the Open Science Foundation web site (https://osf.io/kq8yz/register/565fb3678c5e4a66b5582f67).

2.5.1. Primary outcome measures

Our goal is to test for changes in IADL-related measures and speed of processing at post-assessment and 12-month follow-up. We assess five different IADL-related measures: (1) text-based fraud detection scenario task; (2) driving-simulator based hazard perception task; (3) self-reported IADL; (4) knowledge about driving based on a quiz from the AARP driver safety course; and 5) knowledge about finances and fraud based on a quiz from the ‘Finances and Fraud’ education course. For measuring speed of processing, Useful Field of View [16] and Digit Symbol Substitution [48] were used.

2.5.2. Secondary outcome measures

The secondary outcome measures include possible changes in: computer proficiency [7] and mobile device proficiency [36], numeracy [12,41], reasoning ability (RAPM; [34]; Letter sets; [18]), memory (HVLT; [10]; RAVLT; [40]), and a suite of IADL tasks, including ATM banking and prescription refilling (UMCFAB; [14]). A questionnaire asking about participant's perception of training effectiveness is also administered at post-test. There is also a survey asking if

participants experienced any fraud or engaged in video games or any activities related to IADL, cognitive, or brain training in the last 12-month. The previous experience survey is administered at baseline and follow-up assessments so that we can track if there is any other exposure to those activities before the trial or during the trial.

2.5.3. Rationale for choosing cognitive measures

The cognitive measures chosen in this trial were based on proximal outcome measures of cognitive abilities used in ACTIVE trial [46]. The ACTIVE trial also used Letter sets [18] to measure reasoning, UFOV [16] and DSST [48] to measure speed of processing, and HVLT [10] and RAVLT [40] to measure memory. We decided to use the same cognitive measures for each cognitive ability (memory, reasoning, & speed of processing) because it would allow us not only to track performance on those cognitive abilities in this trial, but also to compare the results with those of ACTIVE trial. We selected multiple measures for memory and speed of processing because it also enables us to use latent measures for those cognitive abilities. Raven's Advanced Progressive Matrices was additionally selected given its broad ability to predict many different kinds of cognitive activity as well as its low cultural loading [2,15].

2.6. Materials

All participants received a Lenovo Thinkpad, model #X201 notebook computer (screen size: 12.1″ 1280 × 800 resolution), along with a charger, and a 20 ft Ethernet cable. In cases where participants did not have WiFi in their home, but did have an Internet modem, they were instructed to use the Ethernet cable to connect to the Internet during in-lab training. In cases where participants did not have Internet in their home, they were given an ATT MiFi unit (ZTE, model #MF923) with 2 GB of data to start, with a protocol to call the lab to replenish the data as needed.

2.7. Treatment fidelity

To ensure treatment fidelity, several measures were taken, including: (1) pilot testing; (2) consistent training of research assistants (RAs); (3) stepwise protocol created in Qualtrics; (4) logging of unexpected technical events with tasks and participant's devices.

2.7.1. Pilot testing

Prior to beginning data collection, all measures and parallel forms were piloted by over 10 RAs and graduate students, to ensure that there were no issues with the tasks themselves. All interventions were also pilot tested by graduate students, project coordinators, and research assistants. The entire protocol (except for the follow-up visits), including one week of at-home laptop training, was pilot tested with six older adult participants.

2.7.2. Training of research assistants

For a period of two weeks, research assistants working on this project were trained on administering: each of the assessments; each of the interventions; and basic laptop use training. After training, research assistants practiced completing entire assessments (baseline, post, follow-up) with each other to ensure their knowledge of the entire protocol. If research assistants had any questions, they would be answered by project coordinators or graduate students, prior to the beginning of data collection.

2.7.3. Stepwise protocol in Qualtrics

A Qualtrics survey was created to step research assistants through a protocol for each assessment, to be completed on a tablet. At the beginning of this survey, interventionists were asked to enter a start code from the participant's folder. This code ensured that the tasks to be completed by participants that day matched the intended tasks for their assessment. Once this code was entered, all measures followed in their intended order for their specific counterbalancing condition. If an interventionist entered an incorrect code, the survey would prompt them to go get a Project Coordinator for assistance, and stop the flow of the protocol.

2.7.4. Reporting of issues

Any issues encountered throughout the course of the intervention, either with a task or device, were logged in a survey, containing anonymous identifiers for participants, and details of their technical issue (e.g., cannot connect to WiFi, automated Windows Update; file upload issue related to device cache).

2.8. Data and safety monitoring

Given that the current project is a single-site, minimal risk study, oversight of the current study is provided by the Principal Investigator (PI), Dr. Charness, and the co-principal investigator, Dr. Boot. The Principal Investigator (PI) is responsible for ensuring participants' safety on a daily basis. The current study is also been monitored by an independent Safety Officer (SO) who is not directly involved in the study. The SO and the PI met at least twice annually, in-person, to review study progress, data quality, and participants' safety.

2.9. Sample

Participants were recruited using various methods and sources, such as: advertisement in local media, publicly accessible phone lists (e.g., Florida voter registration list); from a list of registered volunteers, maintained by the Institute for Successful Longevity (ISL; https://apps.psy.fsu.edu/isl); and from a list of individuals who have given permission to be contacted for research purposes, maintained by the FSU Site of CREATE (www.create-center.org).

As shown in Fig. 2, 595 individuals received telephone pre-screening. After excluding people who did not pass the pre-screening (n = 157) or who did not show interest in participating (n = 70), 368 individuals became eligible for scheduling a baseline assessment. Specifically, 24 people were ineligible at the pre-screening due to their previous experience with AARP Driver Safety Course (n = 16), Brain HQ (n = 5), or Aptima's Mind Frontiers video game (n = 3). Of the 368 eligible individuals, some showed a lack of interest in participating (n = 109), or did not respond to our voicemails (n = 21), resulting in 238 individuals being enrolled for a baseline assessment. At the baseline assessment, eight individuals chose to discontinue participation, resulting in 230 participants being randomly assigned to one of the intervention conditions upon completion of the baseline assessment. We want to note that we initially registered a plan for 200 participant on ClinicalTrials.gov. But recruitment was proceeding more efficiently than anticipated so we decided to increase sample size to increase power. After randomization into conditions, 28 participants chose to discontinue participation, resulting in 202 participants eligible to return for a post-assessment.

Among the randomized participants, the sample was 58% female (133 out of 230), with age ranging from 64 to 92 years (M = 71.35, SD = 5.33). It should be noted that one participant was recruited at age 64 and turned 65 right after post-training assessment. The participants who volunteered were not racially diverse (82% White) and most of them have a college degree and above (64%). No statistically significant difference in gender, age, race, and education was found among intervention conditions. Baseline demographics by intervention condition are summarized in Table 2. As summarized in Table 3, there were also no significant differences in the primary outcome measures at baseline among the intervention conditions.

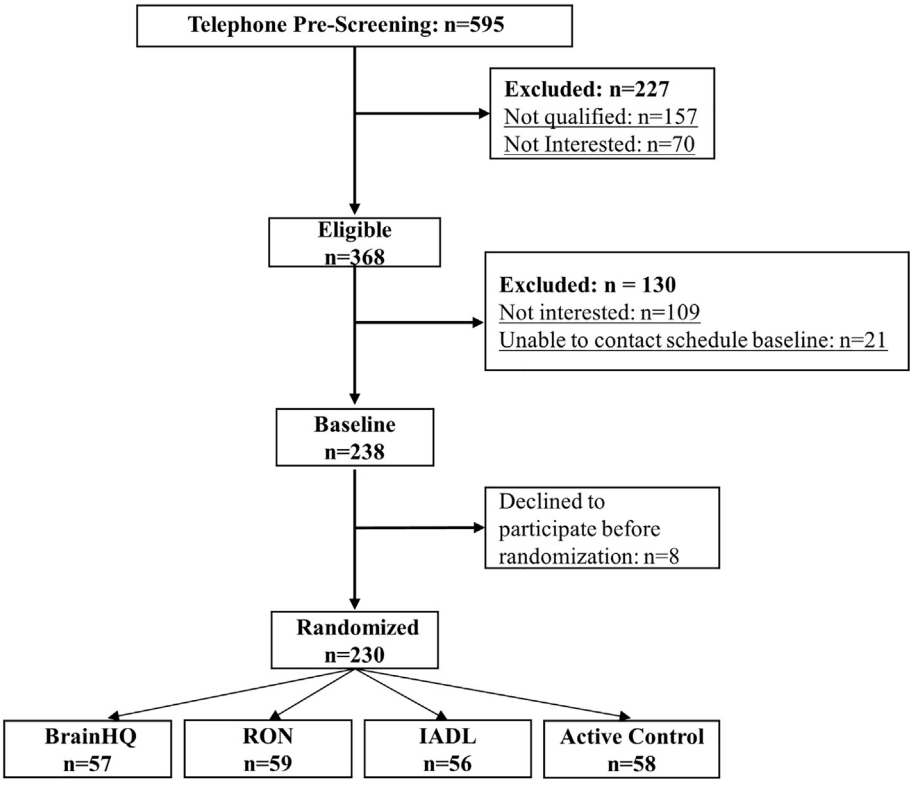

Fig. 2. CONSORT diagram.

3. Discussion

3.1. Project challenges related to technology

3.1.1. Networking challenges

We experienced occasional Internet disconnections which resulted in technical delay for research assistants to use tablets designed to administer assessments. The wireless connection in our lab is shared with the entire department, so at times of high load (e.g., late afternoon), disconnections proved to be problematic. In some instances, this occurred during the audio presentation of stimuli for both the Hopkins Verbal Learning and Rey Auditory Verbal Learning tasks. A potential solution to such interruptions is to ensure that adequate bandwidth is present by pretesting the network under maximal load and adding capacity in advance, budget permitting. Other alternatives include keeping testing resources such as audio files local to each tablet, and using the offline Qualtrics application to display surveys.

3.1.2. Laptop challenges

Due to IT policies with both Microsoft and FSU, automatic updates were not able to be turned off on intervention laptops. As a result, laptops used for training, as well as laptops sent home with the participants would often begin to update themselves, taking upwards of 20 min in most cases. Participants were told to report these occurrences in their diaries. A potential solution is to negotiate exceptions to operating system update policies in the context of research studies.

The laptops used for this intervention were donated by FSU Psychology and College of Medicine's IT representatives, and in some

Table 2
Baseline demographic in the ICE-ACT trial.

	Total (N = 230)	BHQ (N = 57)	RON (N = 59)	IADL (N = 56)	CONTROL (N = 58)
Age	71.4 (SD = 5.3)	71.5 (SD = 5.9)	71.4 (SD = 5.7)	72.4 (SD = 5.5)	70.2 (SD = 3.8)
(64–92)	Median = 70	Median = 70	Median = 70	Median = 71	Median = 69
Gender	58% female (133/230)	58% female (33/57)	56% female (33/59)	64% female (36/56)	53% female (31/58)
Education	HS or less = 8%	HS or less = 7%	HS or less = 5%	HS or less = 13%	HS or less = 9%
	Some college = 28%	Some college = 28%	Some college = 24%	Some college = 27%	Some college = 33%
	Bachelor = 27%	Bachelor = 21%	Bachelor = 31%	Bachelor = 36%	Bachelor = 22%
	Master = 24%	Master = 26%	Master = 32%	Master = 14%	Master = 24%
	Doctoral = 12%	Doctoral = 18%	Doctoral = 8%	Doctoral = 11%	Doctoral = 12%
Race	White = 82%	White = 79%	White = 88%	White = 82%	White = 79%
	Black = 12%	Black = 12%	Black = 7%	Black = 13%	Black = 17%
	Other* = 6%	Other* = 9%	Other* = 5%	Other* = 5%	Other* = 4%
Household	Less than $20 k =10%	Less than $20 k = 14%	Less than $20 k = 8%	Less than $20 k = 5%	Less than $20 k = 12%
Income	$20 k - $39, 999 = 16%	$20 k - $39,999 = 11%	$20 k - $39,999 = 14%	$20 k - $39,999 = 23%	$20 k - $39,999 = 17%
	$40 k - $59,999 = 20%	$40 k - $59,999 = 18%	$40 k - $59,999 = 15%	$40 k - $59,999 = 27%	$40 k - $59,999 = 19%
	$60 k - $79,999 = 20%	$60 k - $79,999 = 12%	$60 k - $79,999 = 31%	$60 k - $79,999 = 16%	$60 k - $79,999 = 21%
	$80 k + = 26%	$80 k + = 37%	$80 k + = 20%	$80 k + = 18%	$80 k + = 29%
	Do not answer = 8%	Do not answer = 8%	Do not answer = 12%	Do not answer = 11%	Do not answer = 2%

* Asian, Native American, Alaska Native Multi-Racial, no primary group, etc.

Table 3
Baseline descriptive of outcome measures (Mean, SD in parentheses) in ICE-ACT trial.

Measure	Measure scale	Total (N = 230)	BHQ (N = 57)	RON (N = 59)	IADL (*N* = 56)	CONTROL (N = 58)	F (p) values
UFOV*	milliseconds						
Processing speed		20.8 (39.5)	25.7 (45.9[3])	25.0 (63.4[3])	16.2 (3.8)	16.1 (4.6)	1.04 (0.38)
Divided Attention		56.3 (86.2)	64.6 (102.3)	47.2 (93.2)	59.0 (65.2)	54.9 (80.2)	0.42 (0.74)
Selective Attention		152.7 (86.7)	156.0 (97.0)	149.3 (90.)	159.3 (84.7)	146.4 (75.4)	0.26 (0.85)
Total score		229.8 (186.3)	246.3 (225.1)	221.6 (218.0)	234.5 (138.5)	217.4 (149.6)	0.28 (0.84)
Digit Symbol	# of correctly						
Substitution*	drawn symbol	44.7 (12.4)	44.9 (13.4)	45.6 (11.9)	45.2 (10.7)	43.3 (13.4)	0.36 (0.78)
Ravens Advanced	# of correct						
Progressive Matrices[1]	response	4.4 (2.3)	4.5 (2.5)	4.7 (2.2)	4.3 (2.2)	4.3 (2.4)	0.28 (0.84)
Letter Sets[1]	# of correctly						
	marked	9.2 (2.8)	9.5 (2.6)	9.1 (2.6)	9.1 (2.9)	9.0 (2.9)	0.34 (0.79)
Hopkins Verbal	# of correctly						
Learning (delayed)	recalled	8.5 (2.5)	8.4 (2.2)	8.6 (2.6)	8.4 (2.8)	8.7 (2.4)	0.20 (0.90)
Rey Auditory Verbal	# of correctly						
Learning (delayed)	recalled	6.0 (3.1)	5.6 (2.9)	6.2 (3.3)	6.0 (3.1)	6.4 (3.0)	0.72 (0.54)
Numeracy	# of correct						
	item	2.5 (1.5)	2.4 (1.5)	2.6 (1.4)	2.5 (1.5)	2.4 (1.5)	0.29 (0.84)
Self-reported IADL*	0 (no difficulty) ~						
	~ 3 (severe difficulty)	0.34 (1.1)	0.32 (0.85)	0.20 (0.85)	0.38 (1.5)	0.47 (1.1)	0.61 (0.61)
Fraud Detection Task*	0–100						
Fraudulent cases	(100% confident of	86.6 (14.3)	90.0 (9.8)	87.0 (16.8)	85.3 (11.9)	83.9 (16.8)	1.93 (0.13)
Non-fraudulent cases	fraud attempt)	38.5 (23.0)	39.4 (23.2)	38.2 (24.0)	34.9 (22.8)	41.3 (22.2)	0.77 (0.51)
Simulated IADL (UMCFAB)	a rate measure =						
ATM task	total correct responses	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.04 (0.02)	0.05 (0.02)	1.01 (0.37)
Prescription refill task	/time	0.06 (0.04)	0.07 (0.07)	0.07 (0.04)	0.06 (0.02)	0.06 (0.03)	0.94 (0.42)
Driving Performance							
in HPT Task[2,*]							
Mean velocity	meters per second	12.3 (2.8)	11.5 (3.1)	12.3 (3.5)	12.6 (1.7)	12.7 (2.4)	1.81 (0.15)
Max brake force	0–1 (max depression)	0.96 (0.16)	0.95 (0.19)	0.94 (0.24)	0.99 (0.04)	1.00 (0.02)	1.94 (0.13)
Lane position	- 1 (1 m to left)	−0.37 (0.15)	−0.37 (0.17)	−0.36 (0.14)	−0.41 (0.12)	−0.37 (0.16)	1.01 (0.39)
	~ (1 m to right)						
CPQ total	6–30 ("very easily"	25.5 (4.4)	25.9 (4.2)	25.7 (3.8)	25.4 (4.5)	24.9 (5.9)	0.52 (0.67)
	across all sub-skills)						
MDPQ total	8–40 ("very easily"	28.4 (9.4)	28.9 (9.5)	28.5 (9.1)	27.4 (9.7)	28.5 (9.6)	0.26 (0.85)
	Across all sub-skills)						

[1] $N = 223$.
[2] $N = 203$.
[3] The SDs were driven by an outlier in each condition: 321.5 in BHQ and 500 in RON. Means (SDs) without them were 20.41 (22.88) and 16.84 (7.72).
* Primary outcome measure. UMCFAB: University of Miami Computer-Based Functional Assessment Battery, HPT: Hazard Perception Task, CPQ: Computer Proficiency Questionnaire, MDPQ: Mobile Device Proficiency Questionnaire.

cases, laptop batteries were at the end of their life, requiring replacement. In these seven cases, participants would either come to the lab to get a replacement or we drove to their house to install a new battery. A potential solution, depending on budget, is to pre-purchase batteries and always start with fresh batteries.

Previous research has shown that double-clicking is sometimes problematic for older adults, particularly choosing the correct double-click speed [29]. For this reason, we turned off double-clicking for opening items, so that participants could open their tasks and diary with a single click. Participants, likely those with double-click experience on their own machines, had difficulties understanding the single-click, leading to numerous tech support calls about multiple tasks open at once (e.g., if double-clicking the diary, two diary modules would open). A potential solution is to advise all participants that a single click is needed and train them on that action.

3.1.3. Training challenges

In general, we encountered the most difficulty training participants for the Rise of Nations video game condition. Most of our participants have not ever played video games, so understanding how to play and administer the game at once was reportedly difficult to comprehend. This game not only had several game elements, and possible victory types, but deep menu structures, necessitating technical support calls for some frustrated participants. A potential solution, depending on budget resources, is to provide more training. As expected, several participants experienced mild to moderate simulator sickness during the Driving Familiarization Task or the Hazard Perception Task. Although the participants were allowed to opt out the simulator tasks, a potential solution to mitigate simulator sickness (e.g., to minimize events possibly causing sickness, such as stopping, starting, and turns in simulator scenarios, to adapt or habituate participants to the simulator over time, pre-screening for normal car sickness) should be considered for future studies with driving simulator tasks.

3.1.4. Intervention challenges

For both the AARP Driver Safety course, and the Financial Fraud training course, Adobe Flash Player was required to be enabled in Google Chrome, even though it was already enabled on the laptop itself. In cases where Flash Player updated, Flash would need to be re-enabled in Chrome, a process we walked participants through on the phone. Because Flash is being abandoned as a supported format by Adobe in 2020, this problem will eventually resolve as users adopt newer standards and formats for creating video-based training materials (e.g., HTML5). However, update and obsolescence problems are unlikely to disappear so technology interventions need to be vigilant about software updates.

For the BrainHQ group, participants reported difficulties understanding the proper sequencing of actions to complete the "Double Decision" task. In this task, participants had to first click which vehicle they saw presented centrally in the display, from set of two options, followed by a click along "slices of a pie" where a 'Route 66' sign appeared. More details can be gathered from the official documentation of

this task: https://www.brainhq.com/why-brainhq/about-the-brainhq-exercises/attention/double-decision. A potential solution is to provide more training on this task.

Although participants were instructed to read the take-home manual, we did not monitor how many times participants consulted with the manual. In particular, for the RON group, participants were instructed to increase the difficulty by one level if they achieved two victories in a row and reduce difficulty by one level if they lost. We did not monitor if participants complied with this rule in real time during the home-intervention. However, some participants made a note of this in the diary and every game session (e.g., lose or win, difficulty level) was backed up to the server, which would allow us to track if participants appropriately adjusted the difficulty level by replaying the session.

3.2. General discussion

Our research is motivated by the need to determine how best to design and administer effective and engaging technology-based cognitive interventions to support independence in adulthood. Following a series of influential Institute of Medicine reports, comparative effectiveness research has become an accepted standard in medical research [45]. Although use of RCT research on cognitive and social support intervention is relatively new to aging [3,13,50], it too could benefit from investigations of comparative effectiveness that are aimed at evaluating the most effective interventions for targeted populations and contexts. In spite of these challenges, the current trial focusing on comparative effectiveness will provide meaningful information on the following question: of the current set of cognitive interventions (many of which are delivered through technology), which yields the greatest return (for improving independence) per hour of time invested?

The trial is also yielding both theoretical and methodological insights concerning the limitations of previous intervention studies, such as: (1) possible practice effects confounding estimates of change scores; (2) lack of an active control condition; and (3) absence of direct measures of independence, relying instead on self-reports of independence. To address such limitations, we used alternate forms not only for cognitive measures but also for the IADL measures. We also included the active control condition and items asking about participants' perception of training effectiveness so that we may check for expectation effects and control for them. We are also attempting to obtain more direct measures of independence in older adults. For driving, we assessed driving simulator performance on a road hazards perception test similar to prior hazard perception tests that are reliably related to driving errors for on-road tests (e.g. [27,42],). We assessed a diverse set of financial outcomes, such as performance on a Fraud Detection task and the simulated banking management task in UMCFAB. A recent U.S. Senate report shows that there is a looming retirement crisis for aging Americans, particularly the baby boom cohorts who have failed to accumulate sufficient wealth to sustain them in retirement [25], perhaps due in part to inadequate financial literacy. By using the newly developed measures in the current trial, future studies can explore effective training techniques for improving financial management and the ability to detect various types of fraud.

Our IADL measures were not without their limitations. We do not have objective and in situ IADL measures, so we had to rely on self-reported measures as well as simulated IADLs. Previous work has shown that self-reported measures of IADLs tend to show range restriction with a large majority of participants scoring near ceiling [9], so we opted to complement these measures with observable, continuous measures of simulated IADL performance where feasible. Hazard perception is an important IADL for predicting crash probability (similar to how UFOV is used) and is typically performed as a reaction-time task on a computer. With this approach there is limited ecological representativeness - when responding to a hazard on the road, we change our behaviors in a multidimensional way (i.e., change acceleration/braking, deviate in lane position). Our approach was to develop a driving simulator based measure of hazard perception, where we are able to explore these multidimensional changes and generate more holistic representations of driving risk. Some have argued that performance on a driving simulator is not a direct analog for “real-world” driving, mainly due to differences in visual, auditory, and tactile stimulation. But recent work has shown that on-road and simulated driving measures can be correlated as high as 0.8–0.97 [19,21].

Although we reported several measures from the fraud detection and the HPT tasks, we will continue to develop better performance indices. For instance, the current study identified and reported several driving simulator measures (e.g., velocity, lane position, max brake force) that reflect comprehensive driving performance. It would be possible to elaborate those measures by linking them with specific types of hazards and/or time windows closer to the onset of hazards. It should also be noted that practice effects may still be evident for some tests even after using alternate forms given that the practice effects vary in the magnitude across cognitive tests [37]. Further, the use of alternate test forms may not entirely eliminate practice effects. Researchers have used alternate forms of a test to minimize practice effects under the assumption that the forms are equivalent. However, non-equivalence of alternate test forms has been reported for some memory measures, including Rey Auditory Verbal Learning Test and the Hopkins Verbal Learning Test [24]. Thus, a different way to score should be considered for those measures (e.g., equipercentile equating) to address the non-equivalence issue in post- and follow-up test analyses [24].

Another caveat from this trial is that the sample was not racially (82% White) or educationally (64% college degree) diverse, which would limit the generalizability of the study results. In fact, the lack of diversity in samples has been a major challenge in cognitive training intervention studies (for a review, see [47]). Specifically, ethnic minority older adults have experienced disparities in research, including cognitive aging research for many reasons: e.g. strict inclusion/exclusion criteria, lack of information about research, mistrust related to historical occurrences of unethical research practices [28]. Further, we recruited some older adults relying on the lists of registered volunteers who have given permission to be contacted for research purposes. In a volunteer study, it is possible that older research participants are more likely to be more highly educated and significantly healthier than non-participants [32]. Similarly, the proportions of White and higher educated population is higher in the list of volunteers than in the local Census report (62% White and 45.5% Bachelor's degree or higher in Leon County, Florida, U.S. retrieved from U.S. Census Bureau, https://www.census.gov/quickfacts/fact/table/leoncountyflorida/EDU685217#EDU685217). Further research should attempt to recruit more diverse populations using more tailored and adaptive recruitment strategies as discussed in previous studies (e.g., [47]).

In conclusion, this trial has relied on technology-based systems to provide a high-quality, standardized behavioral intervention. Our aim was to provide enough detail on methodology so that others may replicate and extend our efforts. Overall, most participants were able to adhere to the intervention schedule (12% attrition at first post-test) without many issues. We have outlined some of the problems that can arise when using such technology-based interventions, and provided some potential solutions going forward.

Funding

The National Institute on Aging/National Institutes of Health supported this work (NIA P01 AG017211, Project CREATE IV – Center for Research and Education on Aging and Technology Enhancement).

Acknowledgements

Neil Charness, Walter R. Boot, Jong Sung Yoon, Ronald Andringa, Erin R. Harrell, and Kat Lewis received support from the National

Institute on Aging Project CREATE IV – Center for Research and Education on Aging and Technology Enhancement (www.create-center.org, NIA P01 AG017211). Nelson A. Roque was also supported by National Institute on Aging Grant T32 AG049676 to The Pennsylvania State University.

References

[1] J.C. Allaire, M. Marsiske, Everyday cognition: Age and intellectual ability correlates, Psychol. Aging 14 (4) (1999) 627–644.
[2] W. Arthur Jr., D.V. Day, Development of a short form for the Raven Advanced Progressive Matrices test, Educ. Psychol. Meas. 54 (1994) 394–403, https://doi.org/10.1177/0013164494054002013.
[3] K. Ball, D.B. Berch, K.F. Helmers, J.B. Jobe, M.D. Leveck, M. Marsiske, ... F.W. Unverzagt, Effects of cognitive training interventions with older adults: a randomized controlled trial, JAMA 288 (18) (2002) 2271–2281.
[4] K. Ball, J.D. Edwards, L.A. Ross, The impact of speed of processing training on cognitive and everyday functions, J. Gerontol. Ser. B 62 (2007) 19–31.
[5] D.E. Barnes, W. Santos-Modesitt, G. Poelke, A.F. Kramer, C. Castro, L.E. Middleton, K. Yaffe, The Mental Activity and eXercise (MAX) trial: a randomized controlled trial to enhance cognitive function in older adults, JAMA Intern. Med. 173 (2013) 797–804.
[6] C. Basak, W.R. Boot, M.W. Voss, A.F. Kramer, Can training in a real-time strategy video game attenuate cognitive decline in older adults? Psychol. Aging 23 (2008) 765–777.
[7] W.R. Boot, N. Charness, S.J. Czaja, J. Sharit, W.A. Rogers, A.D. Fisk, ... S. Nair, Computer proficiency questionnaire: assessing low and high computer proficient seniors, The Gerontologist 55 (3) (2015) 404–411.
[8] W.R. Boot, D.J. Simons, C. Stothart, C. Stutts, The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects, Perspect. Psychol. Sci. 8 (2013) 445–454, https://doi.org/10.1177/1745691613491271.
[9] J.S. Brach, J.M. VanSwearingen, A.B. Newman, A.M. Kriska, Identifying early decline of physical function in community-dwelling older women: performance-based and self-report measures, Phys. Ther. 82 (4) (2002) 320–328.
[10] J. Brandt, The Hopkins Verbal Learning Test: development of a new memory test with six equivalent forms, Clin. Neuropsychol. 5 (1991) 125–142.
[11] M. Calamia, K. Markon, D. Tranel, Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment, Clin. Neuropsychol. 26 (4) (2012) 543–570, https://doi.org/10.1080/13854046.2012.680913.
[12] E.T. Cokely, M. Galesic, E. Schulz, S. Ghazal, R. Garcia-Retamero, Measuring risk literacy: the Berlin Numeracy Test, Judgm. Decis. Mak. 7 (2012) 25–47.
[13] S.J. Czaja, W.R. Boot, N. Charness, W.A. Rogers, J. Sharit, Improving social support for older adults through technology: findings from the PRISM randomized control trial, The Gerontologist 58 (2017), https://doi.org/10.1093/geront/gnw249.
[14] S.J. Czaja, D.A. Loewenstein, S.A. Sabbag, R.E. Curiel, E. Crocco, P.D. Harvey, A novel method for direct assessment of everyday competence among older adults, J. Alzheimers Dis. 57 (4) (2017) 1229–1238.
[15] J. Duncan, M. Schramm, R. Thompson, I. Dumontheil, Task rules, working memory, and fluid intelligence, Psychon. Bull. Rev. 19 (2012) 864–870.
[16] J.D. Edwards, L.A. Ross, V.G. Wadley, O.J. Clay, M. Crowe, D.L. Roenker, K.K. Ball, The useful field of view test: normative data, Arch. Clin. Neuropsychol. 21 (2006) 275–286.
[17] J.D. Edwards, V.G. Wadley, D.E. Vance, K. Wood, D.L. Roenker, K.K. Ball, The impact of speed of processing training on cognitive and everyday performance, Aging Ment. Health 9 (3) (2005) 262–271.
[18] R.B. Ekstrom, J.W. French, H.H. Harman, D. Dermen, Manual for Kit of Factor-referenced Cognitive Tests, Educational Testing Services, Princeton, NJ, 1976.
[19] A. Eriksson, V.A. Banks, N.A. Stanton, Transition to manual: comparing simulator with on-road control transitions, Accid. Anal. Prev. 102 (2017) 227–234.
[20] K.A. Ericsson, R.R. Hoffman, A. Kozbelt, A.M. Williams (Eds.), The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press, 2018.
[21] B. Freund, S. Gravenstein, R. Ferris, E. Shaheen, Evaluating driving performance of cognitively impaired and healthy older adults: a pilot study comparing on-road testing and driving simulation, J. Am. Geriatr. Soc. 50 (2002) 1309–1310.
[22] K.T. Goode, K.K. Ball, M.E. Sloane, D.L. Roenker, D.L. Roth, R.S. Myers, et al., Useful field of view and other neurocognitive indicators of crash risk in older adults, J. Clin. Psychol. Med. Settings 5 (1998) 435–440, https://doi.org/10.1023/A:1026206927686.
[23] C.S. Green, D. Bavelier, Exercising your brain: a review of human brain plasticity and training-induced learning, Psychol. Aging 23 (4) (2008) 692.
[24] A.L. Gross, S.K. Inouye, G.W. Rebok, J. Brandt, P.K. Crane, J.M. Parisi, ... R.N. Jones, Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time, J. Clin. Exp. Neuropsychol. 34 (2012) 758–772.
[25] C.T. Harkin, The Retirement Crisis and a Plan to Solve It, US Senate Committee on Health, Education, Labor and Pensions, Washington, DC, 2012.
[26] C. Hertzog, A.F. Kramer, R.S. Wilson, U. Lindenberger, Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol. Sci. Public Interest 9 (1) (2008) 1–65.
[27] M.S. Horswill, C.N. Kemala, M. Wetton, C.T. Scialfa, N.A. Pachana, Improving older drivers' hazard perception ability, Psychol. Aging 25 (2) (2010) 464–469.
[28] T.B. Hughes, V.R. Varma, C. Pettigrew, M.S. Albert, African Americans and clinical research: evidence concerning barriers and facilitators to participation and recruitment recommendations, The Gerontologist 57 (2015) 348–358.
[29] C.L. Kelley, N. Charness, Issues in training older adults to use computers, Behav. Inform. Technol. 14 (2) (1995) 107–120.
[30] M.P. Lawton, E.M. Brody, Instrumental activities of daily living scale (IADL), The Gerontologist 9 (1969) 179–186.
[31] I. Legault, R. Allard, J. Faubert, Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking, Front. Psychol. 4 (2013) 323.
[32] J.J. Manly, D. Mungas, JGPS special series on race, ethnicity, life experiences, and cognitive aging, J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. 70 (2015) 509–511, https://doi.org/10.1093/geronb/gbv030.
[33] R. Mata, T. Pachur, B. Von Helversen, R. Hertwig, J. Rieskamp, L. Schooler, Ecological rationality: a framework for understanding and aiding the aging decision maker, Front. Neurosci. 6 (2012) 19.
[34] J. Raven, J.C. Raven, J.H. Court, Raven Manual: Section 4, Advanced Progressive Matrices, Oxford Psychologists Press, Oxford, England, 1998.
[35] D.L. Roenker, G.M. Cissell, K.K. Ball, V.G. Wadley, J.D. Edwards, Speed-of-processing and driving simulator training result in improved driving performance, Hum. Factors 45 (2) (2003) 218–233.
[36] N.A. Roque, W.R. Boot, A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire, J. Appl. Gerontol. 37 (2) (2018) 131–156.
[37] T.A. Salthouse, Influence of age on practice effects in longitudinal neurocognitive change, Neuropsychology 24 (2010) 563.
[38] T.A. Salthouse, D.H. Schroeder, E. Ferrer, Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age, Dev. Psychol. 40 (2004) 813.
[39] T.A. Salthouse, E.M. Tucker-Drob, Implications of short-term retest effects for the interpretation of longitudinal change, Neuropsychology 22 (2008) 800.
[40] M. Schmidt, Rey Auditory and Verbal Learning Test: A Handbook, Western Psychological Services, Los Angeles, CA, 1996.
[41] L.M.L. Schwartz, S.S. Woloshin, W.C.W. Black, H.G.H. Welch, The role of numeracy in understanding the benefit of screening mammography, Ann. Intern. Med. 127 (1997) 966–972.
[42] C.T. Scialfa, M.C. Deschênes, J.D. Ference, J. Boone, M.S. Horswill, M. Wetton, Hazard perception in older drivers, Int. J. Hum. Fact. Ergonom. 1 (3) (2012) 221–233.
[43] D.J. Simons, W.R. Boot, N. Charness, S.E. Gathercole, C.F. Chabris, D.Z. Hambrick, E.A. Stine-Morrow, Do “brain-training” programs work? Psychol. Sci. Public Interest 17 (3) (2016) 103–186.
[44] D.J. Souders, W.R. Boot, K. Blocker, T. Vitale, N.A. Roque, N. Charness, Evidence for narrow transfer after short-term cognitive training in older adults, Front. Aging Neurosci. 9 (2017) 41.
[45] H.C. Sox, S. Greenfield, Comparative effectiveness research: a report from the Institute of Medicine, Ann. Intern. Med. 151 (3) (2009) 203–205.
[46] S.L. Tennstedt, F.W. Unverzagt, The ACTIVE study: study overview and major findings, J. Aging Health 25 (2013) 3S–20S.
[47] M. Tzuang, J.T. Owusu, A.P. Spira, M.S. Albert, G.W. Rebok, Cognitive training for ethnic minority older adults in the United States: a review, The Gerontologist 58 (2017) e311–e324.
[48] D. Wechsler, Manual for Wechsler Memory Scaled—Revised, The Psychological Corporation, New York, 1981.
[49] D. Wechsler, Wechsler Adult Intelligence Scale III, 3rd ed., The Psychological Corporation, San Antonio, TX, 1997.
[50] S.L. Willis, S.L. Tennstedt, M. Marsiske, K. Ball, J. Elias, K.M. Koepke, ... E. Wright, Long-term effects of cognitive training on everyday functional outcomes in older adults, JAMA 296 (23) (2006) 2805–2814.