



Published in final edited form as:

Psychol Aging. 2022 March ; 37(2): 210–221. doi:10.1037/pag0000663.

Relative Effectiveness of General versus Specific Cognitive Training for Aging Adults

Nicholas Gray^{1,*}, Jong-Sung Yoon^{2,*}, Neil Charness^{1,*}, Walter R. Boot¹, Nelson A. Roque³, Ronald Andringa¹, Erin R. Harrell⁴, Katharine G. Lewis¹, Thomas Vitale¹

¹Department of Psychology, Florida State University

²Department of Psychology, University of South Dakota

³Department of Psychology, University of Central Florida

⁴Department of Psychology, University of Alabama, Tuscaloosa

Abstract

In the current study we examined three experimental cognitive interventions, two targeted at training general cognitive abilities and one targeted at training specific IADL abilities, along with one active control group to compare benefits of these interventions beyond expectation effects, in a group of older adults (N = 230). Those engaged in general training did so with either the web-based brain game suite Brain HQ or the strategy video game Rise of Nations, while those trained on IADL skills completed instructional programs on driving and fraud awareness. Active control participants completed sets of puzzles. Comparing baseline and post-intervention data across conditions, none of the pre-registered primary outcome measures demonstrated a significant interaction between session and intervention condition, indicating no differential benefits. Analysis of expectation effects showed differences between intervention groups consistent with the type of training. Those in the IADL training condition did demonstrate superior knowledge for specific trained information (driving and finances). Twelve months after training, significant interactions between session and intervention were present in the primary measure of fraud detection, as well as the secondary measures of the letter sets task and Rey's Auditory Verbal Learning Test. However, the specific source of these interactions was difficult to discern. At one-year follow-up those in the IADL condition did not maintain superior knowledge of driving and finances gained through training, as was present immediately post-intervention. Hence, the interventions, when compared to an active control condition, failed to show general or specific transfer in a meaningful or consistent way.

Keywords

cognitive intervention; brain training; instrumental activities of daily living (IADL); older adults; active control; driving; fraud; cognitive ability; RCT

Correspondence concerning this article should be addressed to Dr. Neil Charness at charness@psy.fsu.edu or Department of Psychology, 1107 W. Call Street, Tallahassee, FL 32306-4301.

*These authors contributed equally to this work

Although there is general agreement in the cognitive aging literature that cognitive abilities decline over time, there is significant disagreement about the effectiveness of interventions to counter those declines. Declines are observed in both cross-sectional and longitudinal aging studies with scores on measures of fluid abilities decreasing over the lifespan on the order of 1.5 to 2 standard deviations from the decade of the 20s to the 70s and beyond (e.g., Li et al., 2004; Salthouse, 2010a). Early studies examined whether such declines might be compensated for by acquired skills (e.g., in typing: Salthouse, 1984; in chess playing: Charness, 1981) or even reversed with appropriate interventions (e.g., by training the use of mnemonics, Kliegl, Smith & Baltes, 1989).

Large-scale interventions to try to improve cognition in aging adults began in earnest about 20 years ago with the ACTIVE trial (Jobe et al., 2001; Ball et al., 2002). ACTIVE compared the effectiveness of around 10 hrs of training for specific abilities such as reasoning, memory, and processing speed and assessed near transfer to tasks assessing the trained ability as well as far transfer to tasks of everyday functioning. It was assumed that counteracting decline at the ability level might enable people to live independently longer and possibly forestall or prevent the effects of Alzheimer's disease and related dementias. Potential promise for that approach was seen in the strong relationships observed between cognitive ability measures and measures of everyday activity performance (Allaire & Marsiske, 1999), though both were psychometric measures rather than representative tasks from daily life given the time and difficulty of scoring such activities, e.g., preparing a meal from ingredients available in a real kitchen (Neistadt, 1994).

The idea that general ability training might transfer to specific everyday activity skills was not born out immediately in ACTIVE trial. Instead, as expected, specific transfer was generally found shortly after training. However, testing years later provided some evidence of general transfer relative to the passive control group, including to self-rated difficulties with IADLs (Rebok et al., 2014) and decreases in at-fault vehicle crashes, though paradoxically accompanied by increases in not-at-fault crashes (Ball et al., 2010). Benefits associated with some ACTIVE interventions (e.g., reduced dementia development: Edwards, Xu, Clark, Guey, Ross... et al., 2017; improved health-related quality of life: Wolinsky et al., 2006) years later are somewhat more difficult to attribute to specific interventions given the loss of randomization following booster training sessions and the number and exploratory nature of some analyses which may have capitalized on chance. Nonetheless, the ACTIVE trial was unique in following up participants over an extended time frame (10 years) and providing evidence suggestive of delayed effects for a modest amount of training.

However, there was a strong theoretical precedent for a specific transfer finding from the earliest educational literature (Thorndike, 1924) and the expertise literature (Ericsson, Hoffman, Kozbelt & Williams, 2018), suggesting that acquired skills such as reading, arithmetic, and chess are relatively domain specific with little transfer between them even early in life when the brain is highly plastic (e.g., Sala & Gobet, 2017). Neuroscience approaches to transfer suggest that instead of the principle of identical elements proposed by Thorndike, or the updated model by Taatgen (2013) that includes production rules rather than elements, transfer might be expected based on potential structural and functional

changes of specific brain regions that are common to tasks (Nguyen, Murphy, & Andrews, 2019).

A review of the general “brain training” literature suggested that evidence was weak that training on an ability led to far transfer and more specifically to improvements in everyday activities in older adults (Simons et al., 2016). However, a recent meta-analysis suggests that there are significant improvements in far transfer tasks, on the order of $d=.22$, for training of healthy older adults and those with cognitive impairment (Basak, Qin & O’Connell, 2020), though secondary meta-analysis suggests that transfer effects are generally very small or non-existent (Sala et al., 2019). Further, Moreau (2020) has provided some evidence, particularly for cognitive intervention meta-analyses, that estimates based on assumptions of a unimodal distribution for effect sizes may be misleading, and that mixture modeling may uncover multimodal distributions for effect sizes. That is consistent with the framework proposed for understanding conflicting findings for the benefits of cognitive training by Smid, Karbach and Steinbeis (2020), that emphasizes the role that individual differences might play in responsiveness to training, such as differences in baseline ability, age, motivation, personality, and genetic predispositions. Consistent with that argument, a very recent meta-correlation analysis by Traut, Guild, and Munakata (2021) provides evidence that inconsistent results within the cognitive training literature might be attributable to gains being correlated specifically with initial performance, with low ability persons gaining more from training.

What seems replicable from virtually all meta-analyses is that training on a cognitive task results in improvement for that task. Hence, we would expect that training on an activity that is central to independent living may be expected to benefit that activity more than training on a general cognitive ability, at least in the short term. Training of general cognitive abilities through education (Lövdén, Fratiglioni, Glymour, Lindenberger, & Tucker-Drob, 2020) can have long lasting effects on human performance, perhaps because the hours of investment are several orders of magnitude greater than typical short-term cognitive interventions (e.g., that are typically in the 10 hr range, as seen in ACTIVE). However, the effects of educational training seem to influence the level of attained cognitive ability in early adulthood, with minimal effects on rates of cognitive decline later in life (intercept but not slope of age decline). Also, early in the life span, even moderate-length interventions lasting a year, despite having been shown to have immediate medium to large effect sizes, often experience “fade-out” after a year (Bailey, Duncan, Cunha, Foorman, & Yeager, 2020). So, from a cost-benefit perspective, it may be better to train for near-term benefits that minimize catastrophic failures (vehicle crash, losing financial resources to fraud) rather than hope to show persisting general ability benefits with unknown levels of transfer to everyday activities.

There are also several methodological limitations to previous studies that have tested effectiveness of general ability training on independence in older adults. First, aging researchers tend to rely mainly on cognitive ability proxies to assess independence in everyday activities rather than direct measures of the everyday activities. For instance, many studies have measured performance of the Useful Field of View (UFOV) test instead of using performance data from real or simulated driving to predict driving crashes (Goode

et al., 1998). Second, cognitive measures that are repeated from baseline to post-training sessions can make it difficult to estimate true change due to potential practice/retest effects (Rabbitt, Diggle, Holland, & McInnes, 2004; Salthouse & Tucker-Drob, 2008; Salthouse, 2010b) although increased participant age tends to minimize the practice effects as do alternate forms (Calamia et al., 2012). Lastly, expectation and/or placebo effects have been a methodological threat to the design and interpretation of intervention studies (Boot et al., 2013).

In our study, we chose to focus on two requirements for living independently in much of the United States: safe driving and prudent financial management (fraud avoidance), two instrumental activities of daily living (IADLs; e.g., Lawton & Brody, 1969). We contrasted specific training in those areas with general cognitive training to evaluate the relative effectiveness of different types of training. For general training we chose two cognitive training tasks that have shown efficacy in improving cognitive abilities: BrainHQ, a web-based commercial program based on speed of processing training developed in ACTIVE (Smith et al., 2009), and Rise of Nations, a strategy videogame shown to improve executive functioning (Basak, Boot, Voss, & Kramer, 2008). For specific training, we used a combination of a newly designed tutorial on finances and fraud (described in Yoon et al., 2019) and AARP's web-based training program for older adults aimed at improving knowledge and strategies for driving. That condition was called the IADL training condition. We planned to explore how much improvement participants experienced per unit time invested to begin to evaluate the cost-benefit for types of training. We used a puzzles condition as an active control training condition. This condition has been shown to control for expectation of improvement (Boot et al., 2016), while not demonstrating improvement on cognitive measures for short-term training relative to a training suite that exercised cognitive ability measures in a game-like environment (Souders et al., 2017). We assessed potential improvement for general and specific training conditions with traditional measures of cognitive ability, self-ratings of ability during activities of daily living, as well as new measures of daily living competence: hazard perception during simulator driving and fraud assessment for fraud and non-fraud vignettes. We conducted assessments at baseline, at an immediate post-test, and at a one-year follow-up. Details of the study design and baseline results are given in Yoon et al. (2019). To guard against capitalization on chance for the many possible comparisons across training groups and outcome measures, we registered our analysis plan. For describing study results we adopt the following conventions. Per registration on clinicaltrials.gov, we initially defined primary and secondary outcome variables. We consider statistical tests consistent with our pre-registered analyses as confirmatory tests, and all others as exploratory tests.

In the analysis plan, hypotheses about the comparative advantage of training type (general versus specific) were to be tested by analysis of variance tests for interactions, specifically, measurement occasion (baseline to post-test, baseline to follow-up) by training group (BrainHQ, Rise of Nations, IADL training, puzzle control), followed by post-hoc comparisons. We expected that specific training, the IADL condition, would generate better performance than general training on IADL tasks relative to the puzzle control (H1). We expected that general training, BrainHQ and Rise of Nations, would generate better performance than IADL training on cognitive ability tests relative to the puzzle control

(H2). We expected that the use of alternate forms and a control condition that had strong expectations of improvement would minimize practice and placebo effects (H3), diminishing the size of prior benefits observed for general training relative to control (reduced effect sizes relative to earlier findings).

Method

Transparency and Openness

The summary data presented here are also available preregistered design and analysis plan, on [ClinicalTrials.gov](https://clinicaltrials.gov) (Identifier#: [NCT03141281](https://clinicaltrials.gov/ct2/show/study/NCT03141281)). Detailed procedures and deidentified data are available on the Open Science Framework website at the time of publication.

Design and Participants

Intervention Comparative Effectiveness for Adult Cognitive Training (ICE-ACT) was a single-site randomized and controlled clinical trial with 4-group design including 3 intervention groups (brain training, video game training, or a directed IADL training) and an active control group with puzzle training (see Yoon et al., 2019 for details). This trial was conducted from October, 2017 – through August 2019 with older adults aged 65 years living in the Leon County area (Florida, USA). Participants were required to 1) have a valid driver's license and drive at least once a month; 2) speak English; and 3) have adequate cognitive ability assessed by the Weschler Memory Scale III (story A score of 6 or story B score of 4 if they failed story A). 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 previously experienced any of the intervention tasks used in the current trial (for more detail about ineligibility criteria, see Yoon et al., 2019). From 595 individuals contacted for pre-screening, 230 were randomized and the sample was more likely to be female (58%), with age ranging from 64¹ to 92 years ($M=71.35$, $SD=5.33$). Participants were less racially diverse (82% White) than the US population as a whole (76% White: <https://www.census.gov/quickfacts/fact/table/US/PST045219>) and most had a college degree and above (64% vs. 31% with bachelor's degree in the US: <https://www.census.gov/quickfacts/fact/table/US/PST045219>). No statistically significant difference in gender, age, race, and education was found among the intervention conditions. Outcome assessments were conducted immediately (i.e., baseline), after 4-weeks intervention (i.e., post-training assessment), and 1 year after the intervention (i.e., follow-up). To address and trace potential practice/retest effects we used alternate forms for most of the cognitive and IADL outcome measures across assessments (see Yoon et al., 2019 for a full list of outcome measures that used alternate forms) and measured a perception

¹One participant was recruited at age 64 and turned 65 right after post-training assessment.

of training effectiveness at post-training. The order of alternate forms was counterbalanced across participants to control for any potential differences in difficulty. Below is the list of measures that used alternate forms across assessments: the Fraud Detection Task [baseline (B), immediate post-assessment (P), 12-month follow-up (F)], the Driving Simulator Hazard Perception Task (B, P, F), IADL Training Knowledge - Driving (P, F), IADL Training Knowledge - Finances and Fraud (P, F), a Numeracy Test (B, P, F), Letter Sets (B, P, F), the Hopkins Verbal Learning Test (B, P, F), the Rey Auditory Verbal Learning Test (B, P, F), Raven's Advanced Progressive Matrices (B, P, F). The overall study design and contact schedule is summarized in Figure 1 as a CONSORT Diagram. The trial was conducted based on highly standardized protocols (outlined in Yoon et al., 2019) for recruitment, screening, assessment, intervention administration (including manuals sent home with participants), and data transfer from intervention devices, which were approved by the Institutional Review Board (IRB) at Florida State University.

Interventions

Four different training conditions were designed to compare the effectiveness of general and specific technology-based training not only on basic perceptual and cognitive abilities but also on the performance of simulated tasks of daily living. After baseline assessment and randomization, participants received in-lab training designed to encourage familiarity with how to operate a laptop and the necessary components for their assigned intervention task described below. Participants were then asked to train on their assigned condition for four weeks at home (20 hours) by completing two training sessions a day for a minimum of 30 min per training session. Each session followed the previous training session's adaptive difficulty.

Participants in a computerized cognitive training condition received three different tasks of brain training exercises from the BrainHQ platform developed by Posit Science (<https://www.brainhq.com>). The three tasks focusing on improving speed of processing were Double Decision, Freeze Frame, and Target Tracker. The tasks are somewhat analogous to laboratory assessment and training tasks for speed of processing: 1) Double Decision analogous to UFOV, 2) Freeze Frame analogous to a choice reaction time task, and 3) Target Tracker analogous to the dynamic multiple object tracking task (Legault et al., 2013). We were interested in whether such tasks might show far transfer by way of improved performance on daily living tests (e.g., fraud detection, hazard perception in a driving simulator). Participants in a video game training condition were asked to play a real-time strategy game: the Rise of Nations (RON). In RON, competing against other computer-controlled players, gamers are asked to lead a nation from the ancient age to the information age (see https://en.wikipedia.org/wiki/Rise_of_Nations for details). Participants in a directed IADL training were asked to complete two online IADL tasks: 1) AARP Driver Safety Course and 2) Finances and Fraud education. The AARP driver safety course (<https://www.aarpdriversafety.org>) was designed to refresh an older adults' driving skills and knowledge of safe driving strategies. The Finances and Fraud education module was created in collaboration with an instructional design professional and used a common e-learning platform (i.e., Articulate Storyline2; <https://articulate.com/>) 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. The order of IADL tasks was counterbalanced across participants in this group. Because each IADL online course took around three to five hours to complete, participants completed each course twice during the four-weeks of home intervention to make the total number of training hours approximately equivalent to the other intervention conditions, as training time is a potentially critical confound across conditions. This repetition of material is somewhat similar to repeatedly playing the same games in BrainHQ, RON, or in the Puzzles conditions. Lastly, participants in an active control condition were asked to play three computerized puzzle games during the four-weeks of intervention: 1) Desktop Crossword game (developed and published by Inertia software), 2) Britannica Sudoku Unlimited game (developed and published by Britannica Games), and 3) Britannica Word Search game (developed and published by Britannica Games).

Measures

Primary Outcomes—The primary goal of this trial is to test near- or far-transfer after intervention on activities of daily living and general cognitive ability (i.e., speed of processing). To assess performance on the activities of daily living, we used four IADL related outcome measures: (1) fraud detection scenario task; (2) driving-simulator based hazard perception task; (3) self-reported IADL questionnaire; (4) IADL training knowledge about driving, finances, and fraud based on quizzes from the IADL training (this test was administered at posttest and follow up only). To assess speed of processing we used Useful Field of View (Edwards et al, 2006) and Digit Symbol Substitution (Wechsler, 1981).

Secondary Outcomes—In addition to the primary outcome measures, we tested possible changes in: (1) technology proficiency (Computer Proficiency Questionnaire; Boot et al., 2015, Mobile Device Proficiency Questionnaire; Roque et al., 2018); (2) numeracy (Cokely et al., 2012; Schwartz et al., 1997); (3) reasoning ability (Raven’s Advanced Progressive Matrices; Raven et al., 1998; letter sets; Ekstrom et al., 1976), (4) memory performance (Hopkins Verbal Learning Test; Brandt, 1991; Rey’s Auditory Verbal Learning Test; Schmidt, 1996); and (5) other IADL simulated tasks (i.e., ATM banking and prescription refilling tasks, also referred to as “Miami IADL tasks,” in the University of Miami Computer-Based Functional Assessment Battery, Czaja et al., 2017). We also assessed participant’s perception of training effectiveness on motor, cognitive, and IADL abilities after intervention. There was also a questionnaire querying participant’s previous experience with fraud, video games, any activities related to IADL, cognitive, or brain training in the last 12- months. The previous experience questionnaire was administered at baseline and follow-up assessments to track if there was any other exposure to those activities before the trial or during the trial. Finally, we had a manipulation check for the IADL training tutorials consisting of six questions that probed specific knowledge of the material presented. Three questions were selected from quiz sessions in the AARP driver safety course and the other three questions were selected from quiz sessions in the Finances and Fraud education course.

A description of the metrics used can be found in Table 1S and 2S of the supplemental materials. For more detailed explanations of and rationale to choose each outcome measure, see Yoon et al. (2019). The measures and analysis plans are also detailed on [ClinicalTrials.gov](https://clinicaltrials.gov) (<https://clinicaltrials.gov/ct2/show/NCT03141281>) and on the Open Science Foundation web site (<https://osf.io/kq8yz/register/565fb3678c5e4a66b5582f67>).

It can be noted that when alternate forms were used, analysis was not conducted to ascertain whether they were psychometrically identical, or parallel, prior to testing. However, all analyses conducted on measures that used alternate forms were also conducted with form order as a covariate. The counterbalancing of form order did not have a significant impact on any of the analyses reported here, at either posttest or follow-up. Performance on alternate forms at baseline was also analyzed independent of intervention condition, and a significant difference was only detected among the alternate forms of non-fraudulent vignettes. However, counterbalancing equated exposure to each form across conditions, and finding no significant effect of counterbalance order as a covariate adds credence to the following results.

Results

Baseline, posttest, and follow up mean performance for each intervention condition can be seen in Tables 1 and 2, showing primary and secondary measures respectively. Correlations between average scores at different timepoints for each measure are provided in Tables 3S and 4S.

Posttest

Primary and secondary measures at baseline were compared against those immediately post-intervention between the four treatment groups in 4x2 mixed measures ANOVAs. We investigated interactions of condition and session, assessing any differential benefit at posttest, as well as main effects of condition or session. Given that IADL post-training knowledge was assessed only after posttest, one-way ANOVAs were used to examine any differential benefits after posttest. Specific F , p , and Cohen's f values can be seen in Tables 5S and 6S.

Primary Outcomes

Fraud Detection.: There was no significant evidence for an interaction between session and treatment condition for identifying fraudulent or non-fraudulent scenarios through text-based descriptions (F 's < 1 ; p 's $> .40$). Also, no main effect was evident between sessions for fraudulent or non-fraudulent scenarios (F 's < 3.7 ; p 's $> .05$) and no main effect between treatment conditions was shown (F 's < 1.8 ; p 's $> .15$).

As an exploratory analysis, d' was calculated, comparing baseline and posttest judgments of fraud (or non-fraud) across intervention conditions. There was no apparent interaction between intervention and session within d' values, $F(3,197) = .39$, $p = .76$, Cohen's $f = 0.08$. There was also no significant effect of session, $F(1,197) = 3.05$, $p = .08$, Cohen's $f = 0.12$, or intervention condition, $F(3,197) = 1.56$, $p = .20$, Cohen's $f = 0.15$.

Driving Hazard Perception.: Driving simulator measures analyzed here included average speed, max brake compression, and standard deviation of lane position. Average speed may indicate levels of comfort or caution while driving. Max brake application, if elevated, may represent a strong deceleration as a result of an inaccurate judgment of speed or distance, that is, a failure to anticipate or mitigate a perceived hazard. The standard deviation of lane position, as a measure of variability in lane position over time, provides a measure of vehicular control (see Souders et al., 2020). Average speed, maximum brake compression, and standard deviation of lane position showed no significant evidence of an interaction between session and treatment condition (F 's < 1.6 ; p 's $> .20$). There was also no significant main effect of session for average speed ($F < .5$; $p > .6$). However, there was a decrease in max brake compression from baseline ($M = .99$) to posttest ($M = .95$) (indicative of better planning for hazards), $F(1,116) = 7.45$, $p = .01$, Cohen's $f = 0.25$, and a decrease in standard deviations of lane position at posttest ($M = 0.53$) compared to baseline ($M = 0.56$), $F(1,116) = 4.13$, $p = .04$, Cohen's $f = 0.19$. No main effect of treatment condition was shown for any simulator measure (F 's < 1.75 ; p 's $> .15$). Note that the sample size in these analyses is smaller ($n=120$) because participants were allowed to opt out from the driving simulator tasks due to simulator sickness. A chi-square test showed that the proportion of participants who opted out from the simulator task did not differ by the intervention condition, $\chi^2(3, N=202) = 1.90$, $p = .59$.

IADL Status.: Self-reported IADL status did not show significant interactions between session and treatment condition, nor were there main effects of session or treatment condition (F 's < 1.60 ; p 's $> .20$). The manipulation check for IADL post-training knowledge revealed a significant effect of treatment condition, $F(3,198) = 20.84$, $p < .001$, Cohen's $f = 0.56$. Using Bonferroni's method for multiple comparisons, post hoc tests confirmed that those in the IADL condition showed more accurate knowledge of test items than every other treatment condition. Thus, the IADL training was effective for improving specific knowledge.

Speed of Processing.: There was no significant evidence of an interaction between session and treatment for the digit symbol task or the UFOV task (F 's < 1.50 ; p 's $> .20$). However, performance on the digit symbol task significantly improved from baseline ($M = 45$) to posttest ($M = 48$), $F(1,198) = 41.43$, $p < .001$, Cohen's $f = 0.45$. No main effect of session was shown for the UFOV task (F 's < 3.00 ; p 's $> .05$) and no effect of treatment condition was shown for either task (F 's < 1.00 ; p 's $> .60$).

In summary, type of training, whether general or specific, did not differentially impact the outcome measures.

Secondary Outcomes—Secondary measures were also compared between treatment groups at baseline and immediately post-intervention. As mentioned in the preregistered analysis plan, participants' scores on the Computer Proficiency Questionnaire (CPQ), Mobile Device Proficiency Questionnaire (MDPQ), letter sets test, Raven's Progressive Matrices, Hopkins Verbal Learning Test, and the Rey Auditory Verbal Learning Test were all transformed into z-scores prior to analyses. Z-scores were calculated using means and standard deviations from all baseline data prior to attrition, which means that the z-score for

a specific condition in a post-training or follow-up assessment also indicates a difference in average score compared to the baseline average across conditions (see Table 2 for averages).

Technology proficiency.: There was no significant evidence of an interaction between session and treatment condition for the CPQ or MDPQ (F 's < 2.50; p 's > .05). Likewise, no significant effect of session was present for either measure (F 's < 3.40; p 's > .05) nor was there an apparent effect of treatment condition on the CPQ or the MDPQ (F 's < 1.10; p 's > .35).

Numeracy and Reasoning.: The Numeracy Task, the letter sets test, and Raven's Progressive Matrices showed no significant interaction between session and condition (F 's < 1.80; p 's > .10). No effect of session was present for Numeracy or Raven's tests (F 's < .20; p 's > .70) but the letter sets test did reveal better performance at post-test (Mean z-score = 0.27) than at baseline (Mean z-score = 0.03), $F(1,190) = 14.14$, $p < .001$, Cohen's $f = 0.27$. Treatment group showed no main effect for Numeracy, letter sets, or Raven's tests (F 's < .80; p 's > .50).

Memory.: The Hopkins Verbal Learning Test and Rey Auditory Verbal Learning Test scores did not show a significant session by treatment interaction (F 's < 1.70; p 's > .15). Also, neither measure showed a main effect of session (F 's < .50; p 's > .50), or treatment condition (F 's < 1.20; p 's > .30).

IADL (ATM banking and prescription refill tasks).: The simulated ATM banking task and prescription refill task did not exhibit evidence of an interaction between session and condition (F 's < 1.60; p 's > .15). However, there was a significant effect of session for both the ATM banking task, $F(1,196) = 10.87$, $p = .001$, Cohen's $f = 0.24$, revealing better performance at post-test (Mean tasks per minute = 3.1) than at baseline (Mean tasks per minute = 2.9), and the prescription refill task, $F(1,195) = 6.81$, $p = .01$, Cohen's $f = 0.19$, also showing better performance at posttest (Mean tasks per minute = 4.5) than baseline (Mean tasks per minute = 3.7). There was not significant evidence of a main effect of treatment condition for either task (F 's < 1.40; p 's > .25).

Correlations among various IADL measures are shown in Table 7S. Self-reported IADL difficulty is most strongly correlated with performance on the ATM and prescription tasks, while driving simulation measures and fraud detection measures each seem to represent distinct metrics of IADL capability. This was confirmed by a factor analysis, which showed three distinct factors with loadings on each factor corresponding to these three groups of correlated tasks.

Perception of Training Effectiveness—An assumption of many cognitive intervention studies is that an active control group will equate for expectations between experimental and control conditions (Boot, Simons, Stothart, & Stutts, 2013; Simons et al., 2016). A series of analyses were conducted to explore this assumption within the current study. At the end of the one-month intervention period, participants were asked to rate whether interventions like the one they experienced could improve vision, reaction time, memory, hand-eye coordination, reasoning, multi-tasking, and IADL performance. Although interventions did

not result in differential objective performance improvements in most cases, interventions were often associated with significantly different expectations (Figure 1S in Supplemental materials). Expectations tended to be greatest for the Brain HQ condition.

For reaction time, expectation scores were entered into an ANOVA with intervention condition as a factor. This analysis revealed a main effect of condition ($F(3,185) = 12.07$, $p < .001$, Cohen's $f = 0.42$) that was driven by greater expectations for improvement in the Brain HQ condition relative to all other conditions (all p values $< .01$, with Bonferroni correction). Similarly, for expectation for hand-eye coordination improvement, a main effect of intervention condition was observed, $F(3,185) = 8.77$, $p < .001$, Cohen's $f = 0.35$. This too was driven by greater expectations in the BHQ condition relative to puzzle game control ($p < .01$) and IADL training ($p < .001$) conditions. Expectations for multitasking improvement also differed as a function of condition, $F(3,185) = 12.18$, $p < .001$, Cohen's $f = 0.28$. Expectations were higher for Brain HQ relative to the IADL and puzzle conditions (p s $< .01$).

RON and IADL training were also associated with greater improvement expectations for certain abilities. The previously mentioned multitasking effect was partly driven by higher expectations for the RON condition relative to the IADL and puzzle conditions (p s $< .001$). Expectations also differed for reasoning ability, $F(3,185) = 6.08$, $p < .01$, Cohen's $f = 0.42$. However, contrary to reaction time and hand-eye coordination, this time Brain HQ was associated with lower expectations compared to the RON ($p < .001$) and the puzzle control group ($p < .05$). A main effect was also observed for IADL expectations, $F(3,185) = 5.95$, $p < .01$, Cohen's $f = 0.28$. Not surprisingly, IADL training produced greater expectations for IADL improvement compared to all other groups ($p < .01$). For memory and vision, no main effects were observed ($p = .078$ and $.063$, respectively).

One-year follow-up: Similar to posttest procedure, and adhering to the pre-registration analysis plan, follow-up scores were compared to those at baseline to investigate interactions, as well as main effects of session and treatment condition. Mean and standard deviation values for primary and secondary measures can be seen in Tables 1 and 2 respectively. Specific F , p , and Cohen's f values can be seen in Tables 5S and 6S.

Primary Outcomes

Fraud Detection.: Twelve months after baseline, a significant interaction between session (baseline vs. follow-up) and treatment group was present in successful fraud detection, $F(3,166) = 2.68$, $p = .05$, Cohen's $f = 0.22$, potentially reflecting a relatively high rise in successful fraud detection among those in the IADL condition and a decrease in successful fraud detection for those in the BHQ condition from baseline to follow-up, as shown in Figure 2S. However, no significant relationships between rates of fraud detection were revealed in Bonferroni corrected pairwise comparisons between conditions. No significant interaction between session and treatment was evident for detection of non-fraudulent vignettes ($F < 1.60$; $p > .20$). No main effect of session was present for detection of fraudulent or non-fraudulent vignettes (F 's $< .70$; p 's $> .40$). Likewise, performance on neither set of vignettes showed a main effect of treatment condition (F 's < 1.70 ; p 's $> .15$).

Once again, d' was calculated as an exploratory analysis to compare baseline and follow-up judgments of fraud. There was a no significant interaction between intervention condition and session for d' values, $F(3,165) = 2.28, p = .08$, Cohen's $f = 0.20$. Similarly, no significant effect of session, $F(1,165) = 1.29, p = .26$, Cohen's $f = 0.09$, or intervention condition, $F(3,165) = 0.88, p = .46$, Cohen's $f = 0.13$, was apparent among d' values.

Driving Hazard Perception.: During driving simulation, average speed, max brake compression, and standard deviation of lane position showed no significant interaction between session and treatment group (F 's < 1.60 ; p 's $> .15$). There was a main effect of session on speed, $F(1,98) = 11.90, p = .001$, Cohen's $f = 0.35$, indicating higher speeds at follow-up ($M = 14$) than at baseline ($M = 13$). Yet, neither max brake compression nor standard deviation of lane position showed a significant effect of session (F 's $< .40$; p 's $> .50$). No significant effect of treatment condition was exhibited by speed, max brake compression, or standard deviation of lane position (F 's < 2.30 ; p 's $> .05$).

IADL Status.: Measures of self-reported IADL status revealed no interaction between session and treatment condition, and no main effect of session or treatment condition (F 's < 1.00 ; p 's $> .35$). IADL post-training knowledge did not show a significant effect of treatment condition, $F(3,166) = 2.23, p = .09$, Cohen's $f = 0.20$. Additionally, using Bonferroni's correction, post hoc tests revealed no significant comparisons between conditions.

Speed of Processing.: The digit symbol task and the UFOV task displayed no significant interaction between session and treatment (F 's < 1.00 ; p 's $> .45$). There was a significant main effect of session for the digit symbol task, $F(1,169) = 14.25, p < .001$, Cohen's $f = 0.29$, indicating better performance at follow-up ($M = 48$) than at baseline ($M = 45$), but there was no apparent main effect of session on the UFOV task ($F < 3.40$; $p > .05$). Treatment condition did not meaningfully affect digit symbol or UFOV performance (F 's $< .50$; p 's $> .70$).

Secondary Outcomes

Technology proficiency.: CPQ and MDPQ scores revealed no significant session by treatment condition interaction (F 's < 2.10 ; p 's $> .10$). CPQ scores did demonstrate a main effect of session, $F(1,166) = 29.06, p < .001$, Cohen's $f = 0.42$, indicating more self-reported capability at follow-up (Mean z-score = 0.26) than at baseline (Mean z-score = 0.06). Similarly, MDPQ scores showed an effect of session, $F(1,166) = 12.77, p < .001$, Cohen's $f = 0.28$, indicating more self-reported capability at follow-up (Mean z-score = 0.23) than at baseline (Mean z-score = 0.09). There was no apparent main effect of treatment group for the CPQ or MDPQ (F 's < 1.90 ; p 's $> .10$).

Numeracy and Reasoning.: Performance on the Numeracy Task and Raven's test did not show significant evidence for an interaction between session and condition (F 's < 1.00 ; p 's $> .40$). Scores on the letter sets task, however, did show a significant interaction between session and condition, $F(3,160) = 2.88, p = .04$, Cohen's $f = 0.23$, which may have been driven by the Rise of Nations training group, which was the lowest performing group at baseline but the highest performing group at follow-up, as shown in Figure 3S. Bonferroni

corrected pairwise comparisons did not reveal any significant relationships between different conditions' performance on the letter sets task. The letter sets task also showed a significant main effect of session, $F(1,160) = 6.51, p = .01$, Cohen's $f = 0.20$, indicating better performance at follow-up (Mean z-score = 0.21) than at baseline (Mean z-score = 0.07). Neither the Numeracy task nor Raven's test showed a similar significant effect of session (F 's $< .70$; p 's $> .40$). The Numeracy task, the letter sets task, and Raven's test indicated no effect of treatment condition (F 's < 1.00 ; p 's $> .50$).

Memory.: Hopkins Verbal Learning Test scores showed no interaction between session and condition ($F < .20$; p 's $> .95$). Performance on the Rey Auditory Verbal Learning Test did show a significant interaction between session and condition, $F(3,169) = 3.64, p = .01$, Cohen's $f = 0.25$, perhaps reflecting the fact that the IADL group was the only treatment condition to show decreased performance at follow-up compared to baseline, as shown in Figure 4S. However, Bonferroni corrected pairwise comparisons revealed no significant relationships between conditions. Hopkins Verbal Learning Test and Rey's Auditory Verbal Learning Test showed no significant effect of session ($F < 3.50$; p 's $> .05$), and no significant effect of treatment condition ($F < 1.10$; p 's $> .35$).

IADL (ATM banking and prescription refill tasks).: Rates of completion for the ATM task and the prescription refill task revealed no significant interaction between session and treatment condition ($F < 1.30$; p 's $> .30$). These tasks also showed no significant main effect of session ($F < 2.60$; p 's $> .10$), and no main effect of treatment group ($F < 1.10$; p 's $> .35$).

Discussion

Consistent findings of cognitive ability decline, particularly in fluid abilities, has spurred cognitive aging researchers and a growing "brain training" industry to investigate ways to slow or reverse such changes both to improve the prospects of maintaining independent living and forestalling severe cognitive impairment and dementia. Although training an ability leads to improvements in that ability in the short term, there is remarkably little evidence of transfer to everyday cognition tasks, suggesting that it may be more efficient to train on specific tasks of everyday living. An important constraint on interpreting prior studies of training effects has been the lack of strong control groups, particularly those that control for expectation/placebo effects. Thus, we evaluated the effects of two types of cognitive training, general (BrainHQ, RON) and specific (IADL driving and finance) in comparison to an active control condition (Puzzles) for a wide range of near and far transfer tasks. We also evaluated the potential of one-time training interventions to generate durable effects, by assessing performance immediately post-training and after a year. We also attempted to minimize retest effects in general by using alternate forms.

Short-term Training Effects

Contrary to prior studies, we found no differential effects (session by treatment interactions) after approximately 20 hr of training over a one-month period at post-test when compared to baseline performance for primary and secondary measures with our pre-registered analysis

plan. Because there was no differential improvement in performance, we did not conduct planned comparative effectiveness analysis, assessing gain per unit time invested.

Interestingly, the analyses of perception of training effectiveness clearly indicate that participants expected different outcomes for different types of training, similar to the findings of Boot, Simons, Stothart, and Stutts (2013). Expectations differences appeared to be driven by the demands of the training tasks, for example, the Rise of Nations game often required participants to manage multiple tasks at the same time and reason about game strategy, while Brain HQ often required players to respond quickly to fast moving stimuli. Results suggest caution – despite careful efforts to control expectations (in our case, the puzzle control group was chosen based on a previous control group that produced equal or higher expectations for improvement compared to a cognitive intervention), expectations may still differ between groups. However, it is also important to note that despite significant expectation differences, few performance differences were noted between groups. This suggests that placebo effects in cognitive intervention studies may be less of a concern than previously thought. This is consistent with a recent study by Vodyanyk, Cochrane, Corriveau, Demko and Green (2021). In four experiments, participants' expectations were manipulated, but no corresponding effects were observed on cognitive outcome measures (unlike Foughi et al., 2016). However, given such conflicting findings, additional research needs to resolve if and when differential expectations translate to differential performance improvement.

Long-term Training Effects

We failed to find robust effects of either general or specific training at a one-year follow up. Although we observed three interactions for the planned baseline to 12-month analysis, it is worth considering that a Bonferroni correction for the 17 interactions tested at follow-up alone, would result in an alpha of .003, which would eliminate the significance of these interactions. However, based on pre-registration of analyses, we examined each of the interactions. IADL training resulted in more successful fraud detection, though not better performance on non-fraud detection. Although that result is consistent with the expectation that targeted training is superior to general ability training, the failure to find a similar advantage for detection of non-fraud scenarios reduces confidence in the efficacy of our training package, as does the d' analysis. We may be observing a differential shift in the criterion, being more liberal in calling a scenario fraudulent. Analysis of beta values did not reveal any interactions between intervention condition and session, nor any main effects of intervention or session. However, those in the IADL condition did show the largest numeric drop in beta values from baseline ($M = .65$) to follow-up ($M = .55$). That same IADL training condition resulted in poorer 12-month performance on the Rey Auditory Verbal Learning Test, though not on the Hopkins Verbal Learning test. We are unable to suggest a mechanism that would result in IADL training worsening memory for list items, and the failure to find the same effect for the Hopkins memory lists suggests that this may represent a Type 1 error. For the reasoning outcome measure, the Rise of Nations strategy game condition showed a greater increase than other conditions in Letter Sets performance, though no interaction was found for Raven's matrices. It is plausible to think that a strategy game

would improve general reasoning performance more than other types of training, but again, the failure to find the same effect with the Raven's task is suggestive of Type 1 error.

Practice or Retest Effects

We found few main effects of session, that is, few re-test effects for our outcome measures. We attribute this in part to the use of alternate forms for most of our tests. Though alternate forms do not always eliminate re-test effects (Salthouse & Tucker-Drob, 2008; Salthouse 2010), particularly in perceptual speed tests, they can serve to minimize re-test effects and potentially eliminate them for the kinds of measures which used alternate forms here. While psychometric similarity of the alternate forms was not established prior to testing, it can be noted that very little evidence of differential improvement between intervention conditions was present even among measures which did not use alternate forms, strengthening the assessment that the present interventions were largely ineffective at creating general or specific transfer.

Caveats

We failed to replicate the finding by Basak, Boot, Voss, and Kramer (2008) that Rise of Nations leads to differential improvements in executive function in older adults. This could be due to some of the major differences between studies: 1) we used a strong (Boot, Simons, Stothart & Stutts, 2013) control group: puzzles, that has been shown to equate for expectation effects about improvement in cognition (Boot, et al., 2016) compared to a no-training, no-contact control group in Basak et al. (although expectation effects were not equated in all domains in the present experiment); 2) training procedures varied in time distribution (20 hr over one month for our participants vs 23.5 hr over 7–8 weeks for theirs), and in location (after initial training in the lab, in homes for our participants and always in the lab for Basak et al.'s participants); 3) we had almost three times as many participants per condition (an average of 57 versus 20 hence more likely to have a precise estimate of effect size; 4) we had only one measure of executive functioning in common: Raven's. There is some evidence that home-based, technology-delivered practice can be effective, with little difference between lab and home-based training (Rebok, Tzuang & Parisi, 2019), and any large-scale (community-level) cognitive intervention will need to be carried out in the home environment rather than in labs. Further, commercial packages like BrainHQ are usually only available online.

Failure to find training effects can mean that training was either ineffective, or inadequate in dose to generate near (or far) transfer effects. Countering that hypothesis, training in IADL activities was sufficient to show that there were robust gains (Cohen's $f=.56$) in specific trained knowledge (main effect ANOVA at post-test showing superior performance by the IADL training group) that dissipated over a year, not unlike fading knowledge of specific course content for a cognitive psychology course (Conway, Cohen & Stanhope, 1991). Also, dosage, 20 hr, was comparable to (e.g., Basak et al., 2008) or about double that provided in prior studies that demonstrated training effects (e.g., ACTIVE: Ball et al., 2002). An important caveat is that our sample was not population-representative as was the ACTIVE sample; our sample mainly consisted of advantaged older adults (e.g., in education level:

Yoon et al., 2019) and it may be more difficult to generate improvements at initially higher levels of cognitive performance.

Another caveat is that alternate forms were not tested for psychometric equivalence. Further, pre-registration of an analysis plan, while helping to guard against degrading significance levels compared to potentially many unregistered comparisons, makes it more difficult to justify alternative analyses which may be superior to the registered ones; here, testing for interactions with latent rather than observed variables might result in more sensitive interaction tests.

In spite of the caveats, among the strengths of our study were: preregistering the study and analysis plan, the use of multiple training conditions and a strong control group, short and long-term (one-year) assessment of performance, use of alternate forms to minimize retest effects, and highly standardized procedures for training and testing. Our work lends support to the claim that general cognitive training does not appear to lead to transfer to everyday tasks (Simons et al., 2016), here measured by simulator driving, self-rated IADL performance, detecting fraud, or the Miami computer-simulated IADL tasks. Specific training also failed to augment performance using those same outcome measures, though specific declarative knowledge about finances and driving was improved and well retained, at least immediately following training. It may be necessary to augment educational interventions (IADL training) with realistic on-road (e.g., Bédard et al., 2008) or in-person fraud training to produce robust effects for driving performance or fraud avoidance. The factor analysis of IADL task performance and self-ratings suggests that everyday activities such as driving, financial management, medication management, and self-reports of everyday task difficulties represent different constructs or skills.

Conclusions

Approximately 20 hours of training, using either training aimed at improving cognitive abilities or training aimed at increasing specific knowledge about driving and finances, did not provide differential benefits on a wide range of specific and general transfer outcome measures. We also failed to find robust general improvements at retest, probably because we used alternate forms for our outcome measures, showing partial support for our third hypothesis (H3). At follow-up one year later, we observed a few differential changes (both losses and gains) that failed to replicate across similar outcome measures, hence suggestive of Type 1 error. Thus, considering post-training and follow-up results, we have found no evidence to support our first two hypotheses (H1 & H2) regarding differential effects of general and specific training.

Such negative findings, along with a recent failure to demonstrate memory and intelligence benefits for new language learning after 55 hr of instruction (Berggren, Nilsson, Brehmer, Schmiedek & Lövdén, 2020) dampen enthusiasm for expecting that it will be easy to boost cognition to promote longer periods of independent living in aging adults, at least for those with above average educational attainment. Whether multi-modal training (Stine-Morrow, Parisi, Morrow, & Park, 2008; Park et al., 2014) might make a greater difference is still an open question, or whether a combination of training types might provide greater efficacy,

such as mixing pharmacological and behavioral interventions (Yesavage et al., 2007). Whether such training would work better for those lower in cognitive ability and educational attainment needs further exploration, particularly in light of recent findings of a significant meta-correlation between initial level and gain in performance (Traut, Guild & Munakata, 2021) but also considering the view that responder analysis carries important risks (Tidwell, Dougherty, Chrabaszcz, Thomas & Mendoza, 2014). Research stretching back at least 100 years suggests that building up skill in a specific area of human performance requires dedicated training in that specific area. One avenue to explore would be preventing specific catastrophic errors that endanger health and wellbeing, by using technology to augment, or substitute for failing abilities (Charness, 2019) rather than trying to rehabilitate higher level cognitive abilities through cognitive training.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by a grant from the National Institute on Aging (awarded to Drs. Neil Charness, Walter Boot, Sara Czaja, Joseph Sharit, and Wendy Rogers), under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE), 4 P01 AG 17211. For detailed explanations of protocol and baseline data, see Yoon et al. (2019). Baseline data were also presented in a poster by Yoon et al. (2018), and participants' perceptions of training were published in a proceedings paper by Andringa, Harell, Dieciuc, and Boot (2019). Our analysis plan was preregistered on [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/NCT03141281) (<https://clinicaltrials.gov/ct2/show/NCT03141281>), and this registration also hosts a full list of measures and summary statistics. A full list of measures and our analysis plan were also preregistered using the Open Science Foundation web site (<https://osf.io/kq8yz/register/565fb3678c5e4a66b5582f67>).

This work serves to temper expectations with respect to the impact of cognitive interventions examined here and their potential to benefit everyday performance. We observed no meaningful or consistent transfer of training resulting from general or specific training interventions relative to an active control group.

References

- Andringa R, Harell ER, Dieciuc M, & Boot WR (2019, July). Older adults' perceptions of video game training in the intervention comparative effectiveness for adult cognitive training (ICE-ACT) clinical trial: An exploratory analysis. *International Conference on Human-Computer Interaction*, 125–134. 10.1007/978-3-030-22015-0_10
- Allaire JC & Marsiske M (1999). Everyday cognition: Age and intellectual ability correlates. *Psychology and Aging*, 14, 627–644. [PubMed: 10632150]
- Bailey DH, Duncan GJ, Cunha F, Foorman BR, & Yeager DS (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. 10.1177/1529100620915848 [PubMed: 33414687]
- Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, Morris JN, Rebok GW, Smith DM, Tennstedt SL, Unverzagt FW, & Willis SL (2002). Effects of cognitive training interventions with older adults: A randomized control trial. *Journal of the American Medical Association*, 288, 2271–2281. [PubMed: 12425704]
- Ball K, Edwards JD, Ross LA, & McGwin G Jr. (2010). Cognitive training decreases risk of motor vehicle crash involvement among older drivers. *Journal of the American Geriatrics Society*, 58, 2107–2113. 10.1111/j.1532-5415.2010.03138.x [PubMed: 21054291]
- Basak C, Boot WR, Voss MW, & Kramer AF (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, 23(4), 765–777. 10.1037/a0013494 [PubMed: 19140648]

- Basak C, Qin S, & O'Connell MA (2020). Differential effects of cognitive training modules in healthy aging and mild cognitive impairment: A comprehensive meta-analysis of randomized controlled trials. *Psychology and Aging*, 35(2), 220–249. 10.1037/pag0000442 [PubMed: 32011155]
- Bédard M, Porter MM, Marshall S, Isherwood I, Riendeau J, Weaver B, Tuokko H, Molnar F & Miller-Polgar J (2008). The combination of two training approaches to improve older adults' driving safety. *Traffic Injury Prevention*, 9(1), 70–76. 10.1080/15389580701670705 [PubMed: 18338298]
- Berggren R, Nilsson J, Brehmer Y, Schmiedek F, & Lövdén M (2020). Foreign language learning in older age does not improve memory or intelligence: Evidence from a randomized controlled study. *Psychology and Aging*, 35(2), 212–219. 10.1037/pag0000439 [PubMed: 32011156]
- Boot WR, Charness N, Czaja SJ, Sharit J, Rogers WA, Fisk AD, ... & Nair S (2015). Computer proficiency questionnaire: assessing low and high computer proficient seniors. *The Gerontologist*, 55(3), 404–411. [PubMed: 24107443]
- Boot WR, Simons DJ, Stothart C, & Stutts C (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, 8, 445–454. doi: 10.1177/1745691613491271. [PubMed: 26173122]
- Boot WR, Souders D, Charness N, Blocker K, Roque N, & Vitale T (2016). The gamification of cognitive training: Older adults' perceptions of and attitudes toward digital game-based interventions. *International Conference on Human Aspects of IT for the Aged Population*, 290–300.
- Brandt J (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, 5, 125–142.
- Calamia M, Markon K, & Tranel D (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, 26(4), 543–570. [PubMed: 22540222]
- Charness N (1981). Aging and skilled problem solving. *Journal of Experimental Psychology: General*, 110, 21–38. [PubMed: 6453184]
- Charness N (2020). A framework for choosing technology interventions to promote successful longevity: Prevent, rehabilitate, augment, substitute (PRAS). *Gerontology*, 66(2), 169–175. [PubMed: 31487742]
- Cokely ET, Galesic M, Schulz E, Ghazal S, & Garcia-Retamero R (2012). Measuring risk literacy: The berlin numeracy test. *Judgment and Decision Making*, 7, 25–47.
- Conway MA, Cohen G, & Stanhope N (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General*, 120(4), 395–409. 10.1037/0096-3445.120.4.395
- Czaja SJ, Loewenstein DA, Sabbag SA, Curiel RE, Crocco E, & Harvey PD (2017). A novel method for direct assessment of everyday competence among older adults. *Journal of Alzheimer's Disease*, 57(4), 1229–1238.
- Edwards JD, Ross LA, Wadley VG, Clay OJ, Crowe M, Roenker DL, Ball KK (2006). The useful field of view test: Normative data. *Archives of Clinical Neuropsychology*, 21, 275–286. [PubMed: 16704918]
- Edwards JD, Xu H, Clark DO, Guey LT, Ross LA, & Unverzagt FW (2017). Speed of processing training results in lower risk of dementia. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3, 603–611. 10.1016/j.trci.2017.09.002 [PubMed: 29201994]
- Ekstrom RB, French JW, Harman HH, & Dermen D (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Services.
- Foroughi CK, Monfort SS, Paczynski M, McKnight PE, & Greenwood PM (2016). Placebo effects in cognitive training. *Proceedings of the national Academy of Sciences*, 113(27), 7470–7474. 10.1073/pnas.1601243113
- Goode KT, Ball KK, Sloane M, Roenker DL, Roth DL, Myers RS, & Owsley C (1998). Useful field of view and other neurocognitive indicators of crash risk in older adults. *Journal of Clinical Psychology in Medical Settings*, 5(4), 425–440.
- Jobe JB, Smith DM, Ball K, Tennstedt SL, Marsiske M, Willis SL, ... & Kleinman, K. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials*, 22(4), 453–479. [PubMed: 11514044]

- Kliegl R, Smith J, & Baltes PB (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, 25, 247–256.
- Lawton MP, & Brody EM (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179–186. [PubMed: 5349366]
- Legault I, Allard R, & Faubert J (2013). Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking. *Frontiers in Psychology*, 4, 323. [PubMed: 23761025]
- Li S–C, Lindenberger U, Hommel B, Aschersleben G, Prinz W, & Baltes PB (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, 15, 155–163. [PubMed: 15016286]
- Lövdén M, Fratiglioni L, Glymour MM, Lindenberger U, & Tucker-Drob EM (2020). Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest*, 21(1), 6–41. 10.1177/1529100620920576 [PubMed: 32772803]
- Moreau D (2021). Shifting minds: A quantitative reappraisal of cognitive-intervention research. *Perspectives on Psychological Science*, 16(1), 148–160. 10.1177/1745691620950696 [PubMed: 33021890]
- Neistadt ME (1994). A meal preparation treatment protocol for adults with brain injury. *American Journal of Occupational Therapy*, 48(5), 431–438. DOI:10.5014/ajot.48.5.431
- Nguyen L, Murphy K, & Andrews G (2019). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. *Psychological Bulletin*, 145(7), 698–733. 10.1037/bul0000196 [PubMed: 30998045]
- Park D, Lodi-Smith J, Drew L, Haber S, Hebrank A, Bischof G, & Aamodt W (2014). The impact of sustained engagement on cognitive function in older adults: The synapse project. *Psychological Science*, 25(1), 103–112. 10.1177/0956797613499592 [PubMed: 24214244]
- Rabbitt P, Diggle P, Holland F, & McInnes L (2004). Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59(2), 84–97.
- Rebok GW, Ball K, Guey LT, Jones RN, Kim H-Y... et al. (2014). Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, 62(1), 16–24. DOI: 10.1111/jgs.12607 [PubMed: 24417410]
- Rebok GW, Tzuang M, & Parisi JM (2019). Comparing web-based and classroom-based memory training for older adults: The ACTIVE memory works™ study. *The Journals of Gerontology: Series B*, 75(6), 1132–1143. 10.1093/geronb/gbz107.
- Sala G, & Gobet F (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, 26, 515–520. 10.1177/2F0963721417712760. [PubMed: 29276344]
- Sala G, Aksayli ND, Tatlidil KS, Tatsumi T, Gondo Y, & Gobet F (2019). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, 5(1), 18. DOI: 10.1525/collabra.203
- Salthouse TA (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, 13, 345–371.
- Salthouse TA, & Tucker-Drob EM (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, 22(6), 800–811. 10.1037/a0013091. [PubMed: 18999354]
- Salthouse TA (2010a). Major issues in cognitive aging. Oxford: Oxford University Press.
- Salthouse TA (2010b). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, 24(5), 563–572. 10.1037/a0019026 [PubMed: 20804244]
- Schmidt M (1996). *Rey auditory and verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services.
- Schwartz LML, Woloshin SS, Black WCW, & Welch HGH (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966–972. [PubMed: 9412301]

- Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, & Stine-Morrow EAL (2016). Do “brain training” programs work? *Psychological Science in the Public Interest*, 17, 108–191. DOI: 10.1177/1529100616661983
- Smid CR, Karbach J, & Steinbeis N (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*. 29, 531–537. 10.1177/0963721420951599
- Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Machncke HW, & Zelinski EM (2009). A cognitive training program based on principles of brain plasticity: Results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. *Journal of the American Geriatrics Society*, 57, 594–603. 10.1111/j.1532-5415.2008.02167.x [PubMed: 19220558]
- Souders DJ, Boot WR, Blocker K, Vitale T, Roque NA, & Charness N (2017). Evidence for narrow transfer after short-term cognitive training in older adults. *Frontiers in Aging Neuroscience*, 9(41), 1–10. doi: 10.3389/fnagi.2017.00041 [PubMed: 28174533]
- Souders DJ, Charness N, Roque NA, & Pham H (2020). Aging: Older adults’ driving behavior using longitudinal and lateral warning systems. *Human factors*, 62(2), 229–248. [PubMed: 31361158]
- Stine-Morrow EAL, Parisi JM, Morrow DG, & Park DC (2008). The effects of an engaged lifestyle on cognitive vitality: A field experiment. *Psychology and Aging*, 23(4), 778–786. doi:10.1037/a0014341 [PubMed: 19140649]
- Taatgen NA (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120, 439–471. DOI: 10.1037/a0033138 [PubMed: 23750831]
- Thorndike EL (1924). Mental discipline in high school studies. *Journal of Educational Psychology*, 15, 1–22. 10.1037/h0075386
- Tidwell JW, Dougherty MR, Chrabaszcz JR, Thomas RP, & Mendoza JL (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review*, 21(3), 620–628. 10.3758/s13423-013-0560-7 [PubMed: 24307249]
- Traut HJ, Guild RM, & Munakata Y (2021). Why does cognitive training yield inconsistent benefits? A meta-analysis of individual differences in baseline cognitive abilities and training outcomes. *Frontiers in Psychology*, 26. 10.3389/fpsyg.2021.662139
- U.S. Census Bureau QuickFacts: United States. (2019, July 1). Census Bureau QuickFacts. <https://www.census.gov/quickfacts/fact/table/US/PST045219>
- Vodyanyk M, Cochrane A, Corriveau A, Demko Z, & Green CS (2021). No evidence for expectation effects in cognitive training tasks. *Journal of Cognitive Enhancement*, 5, 296–310. 10.1007/s41465-021-00207-6
- Wechsler D (1981). *Manual for wWechsler mMemory Sscaled—rRevised*. New York: The Psychological Corporation.
- Wikipedia contributors. (n.d.). Rise of Nations. Wikipedia. Retrieved February 16, 2021, from https://en.wikipedia.org/wiki/Rise_of_Nations
- Wolinsky FD, Unverzagt FW, Smith DW, Jones R, Wright E, & Tennstedt SL (2006). The effects of the ACTIVE cognitive training trial on clinically relevant declines in health-related quality of life. *Journals of Gerontology: Social Sciences*, 61B (5), S281–S287. 10.1093/geronb/61.5.S281
- Yesavage J, Hoblyn J, Friedman L, Mumenthaler M, Schneider B & O'Hara R (2007). Should one use medications in combination with cognitive training? If so, which ones? *Journal of Gerontology: Psychological Sciences*, 62, 11–18. 10.1093/geronb/62.special_issue_1.11
- Yoon J, Roque N, Andringa R, Griffis Lewis K, Harrell E, Vitale T, & Charness N (2018). Intervention Comparative Effectiveness for Adult Cognitive Training (ICE-ACT)(CLINICALTRIALS#: [NCT03141281](https://clinicaltrials.gov/ct2/show/study/NCT03141281)). *Innovation in Aging*, 2(suppl_1), 963–964.
- Yoon J-S, Roque NA, Andringa R, Harrell ER, Lewis KG, Vitale T, Charness N, & Boot WR (2019). Intervention comparative effectiveness for adult cognitive training (ICE-ACT) trial: Rationale, design, and baseline characteristics. *Contemporary Clinical Trials*, 78, 76–87. 10.1016/j.cct.2019.01.014 [PubMed: 30711665]

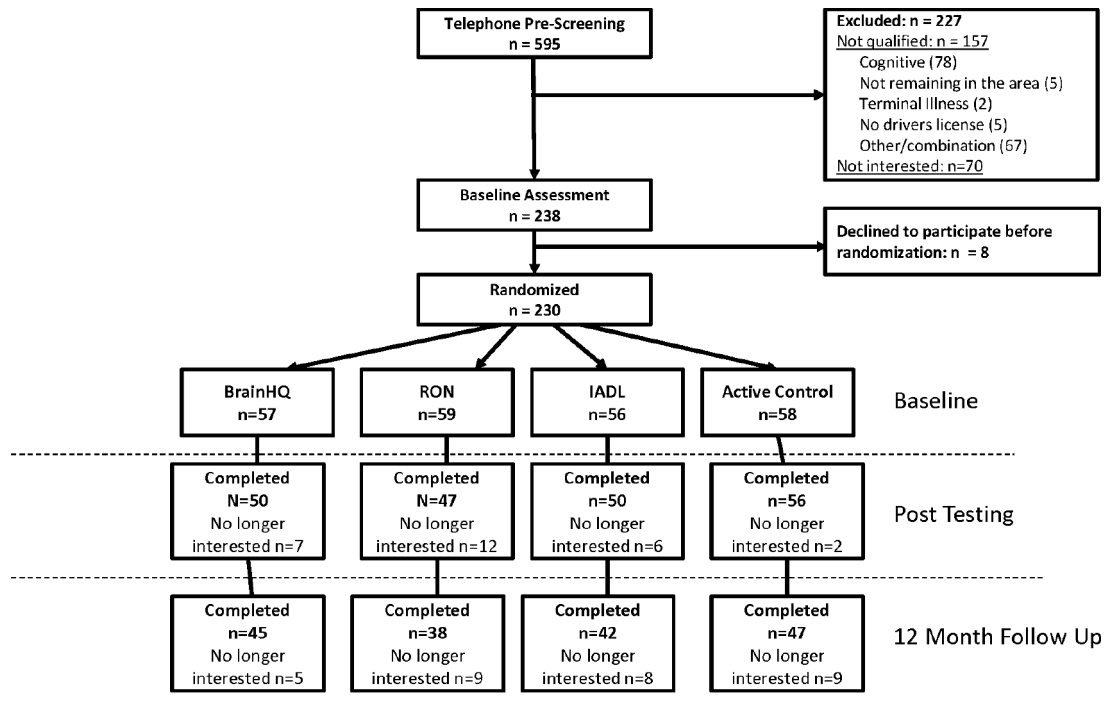


Figure 1.
CONSORT diagram for Comparative Effectiveness Study

Table 1.

Primary Outcome Measures by Intervention and Session

Primary Measure	Primary Measures					
	Brain HQ			Rise of Nations		
	Base	Post	Follow	Base	Post	Follow
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Fraud Detection	90.0 (9.8)	85.9 (13.8)	87.8 (9.0)	85.3 (11.9)	86.3 (10.9)	87.2 (12.9)
Non-Fraud Detection	39.4 (23.2)	32.8 (21.6)	37.7 (23.4)	34.9 (22.8)	36.8 (20.3)	38.3 (22.4)
Driving Simulation						
Speed	11.5 (3.1)	13.3 (1.8)	13.4 (1.5)	12.6 (1.7)	12.9 (4.0)	14.1 (1.0)
Max Brake Compression	0.95 (0.2)	1.0 (0.0)	1 (0.1)	1.0 (0.0)	0.9 (0.3)	1.0 (0.0)
Lane Position	0.54 (0.1)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)	0.5 (0.2)	0.6 (0.1)
Self-Report IADL Status	0.32 (0.9)	0.4 (1.0)	0.4 (1.5)	0.4 (1.5)	0.2 (0.9)	0.2 (0.6)
Digit Symbol	44.9 (13.4)	48.9 (11.9)	48.4 (13.2)	45.2 (10.7)	48.7 (10.4)	47.7 (15.0)
UFOV	246.0 (225.1)	170.0 (194.4)	200.9 (215.8)	234.5 (138.5)	217.5 (192.4)	158.7 (68.7)
Near Transfer - IADL	NA	3.1 (1.0)	2.9 (1.1)	NA	3.2 (1.2)	3.4 (1.0)
Driving and Fraud Avoidance Training						
Puzzle Solving						
	Base	Post	Follow	Base	Post	Follow
Primary Measure	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Fraud Detection	85.3 (11.9)	85.9 (19.4)	91.8 (7.6)	83.9 (16.8)	83.8 (14.9)	85.6 (11.6)
Non-Fraud Detection	34.9 (22.8)	30.2 (24.9)	31.9 (22.4)	41.3 (22.2)	38.7 (20.6)	34.2 (20.7)
Driving Simulation						
Speed	12.6 (1.7)	11.9 (3.9)	13.3 (1.2)	12.7 (2.4)	13.2 (2.5)	13.7 (1.5)
Max Brake Compression	1.0 (0.0)	0.9 (0.3)	1.0 (0.0)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
Lane Position	0.6 (0.1)	0.5 (0.2)	0.6 (0.1)	0.5 (0.1)	0.5 (0.1)	0.5 (0.1)
Self-Report IADL Status	0.4 (1.5)	0.2 (1.4)	0.1 (0.4)	0.5 (1.1)	0.2 (0.7)	0.3 (1.3)
Digit Symbol	45.2 (10.7)	46.8 (12.2)	48.1 (9.8)	43.3 (13.4)	47.1 (12.8)	46.9 (12.3)
UFOV	234.5 (138.5)	210.9 (154.3)	214.7 (152.9)	217.4 (149.6)	219.9 (215.3)	192.2 (148.2)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		<i>Primary Measures</i>					
		Brain HQ		Rise of Nations			
		Base	Post	Follow	Base	Post	Follow
Primary Measure	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Near Transfer - IADL	NA	NA	4.5 (1.3)	3.4 (1.0)	NA	2.9 (1.1)	3.2 (1.0)

Table 2.

Secondary Outcome Measures by Intervention and Session

Secondary Measure	Secondary Measures					
	Brain HQ			Rise of Nations		
	Base	Post	Follow	Base	Post	Follow
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
CPQ (Z-score)	0.11 (0.95)	0.24 (0.9)	0.2 (0.89)	0.06 (0.87)	0.16 (0.95)	0.5 (0.7)
MDPQ (Z-score)	0.08 (1.01)	0.23 (1)	0.15 (1.02)	0.03 (0.97)	0.09 (0.93)	0.39 (0.86)
Numeracy	2.42 (1.53)	2.45 (1.47)	2.4 (1.45)	2.63 (1.38)	2.66 (1.7)	2.76 (1.4)
Letter Sets (Z-score)	0.14 (1.06)	0.24 (0.96)	0.21 (0.99)	-0.01 (0.96)	0.4 (0.84)	0.36 (0.94)
Raven's (Z-score)	0.03 (1.09)	0.08 (1.03)	-0.04 (1.07)	0.11 (0.95)	0.13 (0.97)	0.08 (0.91)
Hopkins (Z-score)	-0.02 (0.87)	0.23 (0.97)	0.19 (1)	0.05 (1)	0.05 (0.95)	0.21 (0.79)
Rey (Z-score)	-0.14 (0.93)	-0.03 (1.04)	0.23 (1.23)	0.05 (1.08)	-0.03 (1)	0.22 (0.89)
IADL - ATM	3.07 (1.31)	3.19 (1.07)	3.09 (1.11)	3.73 (1.01)	3.23 (1.01)	3.26 (1.14)
IADL - Prescription	4.28 (4.21)	4.75 (2.49)	4.22 (2.31)	3.77 (1.69)	4.91 (5.43)	3.98 (1.09)
	Driving and Fraud Avoidance Training					
						Puzzle Solving
	Base	Post	Follow	Base	Post	Follow
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
CPQ (Z-score)	0 (1.01)	0.08 (1.01)	0.29 (0.82)	-0.11 (1.13)	-0.06 (1.2)	0.11 (1.13)
MDPQ (Z-score)	-0.08 (1.03)	-0.04 (1.07)	0.2 (1.02)	0.03 (1.02)	0.07 (1.04)	0.2 (1.07)
Numeracy	2.52 (1.54)	2.16 (1.65)	2.14 (1.81)	2.4 (1.54)	2.48 (1.78)	2.52 (1.69)
Letter Sets (Z-score)	0.01 (1.06)	0.23 (0.97)	0.22 (0.9)	-0.04 (1.05)	0.21 (1.05)	0.06 (1.14)
Raven's (Z-score)	-0.04 (0.95)	-0.01 (1.03)	0.01 (1.1)	-0.03 (1.06)	-0.07 (1.19)	-0.07 (1.09)
Hopkins (Z-score)	-0.02 (1.11)	-0.12 (1.22)	0.21 (1.15)	0.1 (0.94)	0.22 (0.93)	0.4 (0.87)
Rey (Z-score)	0 (1.01)	-0.04 (1.1)	-0.09 (1)	0.12 (0.97)	0.24 (1.07)	0.44 (1.14)
IADL - ATM	2.72 (1.2)	2.98 (1.07)	3.12 (1.26)	2.81 (1.11)	2.97 (1.1)	2.85 (1.04)
IADL - Prescription	3.58 (1.32)	4.2 (1.58)	4.08 (2.13)	3.44 (1.49)	4.14 (3.15)	4.26 (2.04)