ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
2016, Vol. 17(3) 103–186

DOI: 10.1177/1529100616661983
pspi.sagepub.com
SAGE

Do "Brain-Training" Programs Work?

Daniel J. Simons[1], Walter R. Boot[2], Neil Charness[2,3], Susan E. Gathercole[4,5], Christopher F. Chabris[6,7], David Z. Hambrick[8], and Elizabeth A. L. Stine-Morrow[9,10]

[1]Department of Psychology, University of Illinois at Urbana-Champaign; [2]Department of Psychology, Florida State University; [3]Institute for Successful Longevity, Florida State University; [4]Medical Research Council Cognition and Brain Sciences Unit, Cambridge, UK; [5]School of Clinical Medicine, University of Cambridge; [6]Department of Psychology, Union College; [7]Geisinger Health System, Danville, PA; [8]Department of Psychology, Michigan State University; [9]Department of Educational Psychology, University of Illinois at Urbana-Champaign; and [10]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

Summary

In 2014, two groups of scientists published open letters on the efficacy of brain-training interventions, or "brain games," for improving cognition. The first letter, a consensus statement from an international group of more than 70 scientists, claimed that brain games do not provide a scientifically grounded way to improve cognitive functioning or to stave off cognitive decline. Several months later, an international group of 133 scientists and practitioners countered that the literature is replete with demonstrations of the benefits of brain training for a wide variety of cognitive and everyday activities. How could two teams of scientists examine the same literature and come to conflicting "consensus" views about the effectiveness of brain training?

In part, the disagreement might result from different standards used when evaluating the evidence. To date, the field has lacked a comprehensive review of the brain-training literature, one that examines both the quantity and the quality of the evidence according to a well-defined set of best practices. This article provides such a review, focusing exclusively on the use of cognitive tasks or games as a means to enhance performance on other tasks. We specify and justify a set of best practices for such brain-training interventions and then use those standards to evaluate all of the published peer-reviewed intervention studies cited on the websites of leading brain-training companies listed on Cognitive Training Data (www.cognitivetrainingdata.org), the site hosting the open letter from brain-training proponents. These citations presumably represent the evidence that best supports the claims of effectiveness.

Based on this examination, we find extensive evidence that brain-training interventions improve performance on the trained tasks, less evidence that such interventions improve performance on closely related tasks, and little evidence that training enhances performance on distantly related tasks or that training improves everyday cognitive performance. We also find that many of the published intervention studies had major shortcomings in design or analysis that preclude definitive conclusions about the efficacy of training, and that none of the cited studies conformed to all of the best practices we identify as essential to drawing clear conclusions about the benefits of brain training for everyday activities. We conclude with detailed recommendations for scientists, funding agencies, and policymakers that, if adopted, would lead to better evidence regarding the efficacy of brain-training interventions.

Keywords

brain training, cognitive training, learning, transfer, cognitive, skill

Corresponding Author:
Daniel J. Simons, Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, IL 61820
E-mail: dsimons@illinois.edu

Introduction

Spend a few minutes listening to public radio, surfing the Internet, or reading magazines, and you will be bombarded with advertisements touting the power of brain training to improve your life. Lumosity converts basic cognitive tasks into games and has noted in an introductory video that "every game targets an ability important to you, like memory, attention, problem-solving, and more" ("Learn How Lumosity Works" video previously hosted at www.lumosity.com: "Cutting Edge Science Personalized for You," 2015). Posit Science teamed up with the AARP (formerly the American Association of Retired Persons) to offer a version of its BrainHQ software as part of a "Staying Sharp" membership (http://www.aarp.org/ws/miv/staying-sharp/). Cogmed markets its working-memory training program to schools and therapists, claiming that it "will help you academically, socially, and professionally" by "allowing you to focus and resist distractions better" ("How Is Cogmed Different," 2015). And CogniFit has promised to "add useful cognitive training programs for your daily life" ("Improve Your Brain While Having Fun," 2015).

Such statements are standard fare in the marketing materials of brain-training companies, and most back their claims by appealing to the expertise of their founders and/or by citing supporting published research. The aforementioned video emphasizes that Lumosity is "based on neuroscience research from top universities around the world," and elsewhere on the website the company provides a bibliography of 46 papers, posters, and conference presentations from its Human Cognition Project (www.lumosity.com/hcp/research/bibliography). Posit Science's website notes that BrainHQ was "built and tested by an international team of top neuroscientists and other brain experts" and has claimed real benefits shown in "more than 70 published papers" ("Brain Training That Works," 2015), stating that "no other program has this level of proof." Cogmed, too, notes that its program was "developed by leading neuroscientists" and claims that "no other brain-training product or attention-training method can show this degree of research validation" ("Frequently Asked Questions," 2015). CogniFit has promised "fun addictive games designed by neuroscientists" ("Improve Your Brain While Having Fun," 2015).

But does the published research support the claim that such brain-training interventions, or "brain games," improve real-world performance on tasks that matter in our academic, personal, or professional lives? In October 2014, the Stanford Center on Longevity and the Max Planck Institute for Human Development issued an open letter, signed by an international group of more than 70 psychologists and neuroscientists, that "[objected] to the claim that brain games offer consumers a scientifically grounded avenue to reduce or reverse cognitive decline," arguing instead that "there is no compelling scientific evidence to date that they do" ("A Consensus on the Brain Training Industry From the Scientific Community," 2014).

Then, in December 2014, a group of 133 scientists and therapists countered with their own open letter on a website called Cognitive Training Data (www.cognitivetrainingdata.org/), claiming that "a substantial and growing body of evidence shows that certain cognitive-training regimens can significantly improve cognitive function, including in ways that generalize to everyday life." Like the Stanford/Max Planck letter, the response letter concurred that "claims promoting brain games are frequently exaggerated, and are often misleading," but it argued that the literature is replete with "dozens of randomized controlled trials published in peer-reviewed journals that document specific benefits of defined types of cognitive training." The signatories argued that

> many of these studies show improvements that encompass a broad array of cognitive and everyday activities, show gains that persist for a reasonable amount of time, document positive changes in real-life indices of cognitive health, and employ control strategies designed to account for "placebo" effects. (para. 7)

In January 2016, the U.S. Federal Trade Commission (FTC; 2016a) announced that it had charged Lumos Labs with "deceptive advertising" regarding some of the claims the company had made about Lumosity's efficacy and simultaneously announced that the company had agreed to settle the government's $50 million judgment against it by paying a $2 million fine (reduced because of financial hardship) and agreeing to change some of its sales and marketing practices. "Lumosity preyed on consumers' fears about age-related cognitive decline, suggesting their games could stave off memory loss, dementia, and even Alzheimer's disease. But Lumosity simply did not have the science to back up its ads," an FTC official noted (Federal Trade Commission, 2016a). Speaking to *NBC Nightly News*, a staff lawyer for the FTC added, "There just isn't evidence that any of that [using Lumosity] will translate into any benefits in a real-world setting" ("Lumosity to Pay $2M," 2016). The government and Lumos Labs agreed that any future claims of Lumosity's efficacy would have to be backed by "competent and reliable scientific evidence" (Federal Trade Commission, 2016a). The settlement specified that with respect to "performance at school, at work, and in athletics; delaying age-related decline; and reducing cognitive impairment," this standard would require tests that are "randomized, adequately controlled, and blinded to the maximum extent practicable." For other claims, the FTC required evidence of research that is "generally accepted in the profession to yield accurate and reliable results" (FTC, 2016c, pp. 5–7).

How can these conflicting letters, claims, and charges be reconciled? Does cognitive training improve cognitive abilities in ways that generalize beyond the practiced tasks? Do those randomized controlled trials cited in the response letter justify claims about the effectiveness of commercial brain-training products and/or other forms of cognitive training? What skills are improved by cognitive training, and can it improve performance beyond the laboratory?

This article presents a comprehensive review that examines the literature on brain training with the goal of evaluating both the quantity and the quality of the evidence that brain training improves performance on tasks other than the trained ones. In particular, we focus on whether brain-training interventions improve cognition in ways that might improve real-world performance.

What is brain training?

A comprehensive review of all of the pathways to optimal cognitive functioning, some of which might be considered brain training, is beyond our scope (see Hertzog, Kramer, Wilson, & Lindenberger, 2008, for a review of some of these pathways). For example, we do not review evidence for the cognitive benefits of health and fitness interventions (e.g., Hillman, Erickson, & Kramer, 2008; Voss, Vivar, Kramer, & van Praag, 2013), meditation (Gard, Hölzel, & Lazar, 2014; Hölzel et al., 2011), drugs and nutrition (e.g., Burkhalter & Hillman, 2011), strategy training, or lifestyle modification (e.g., Stine-Morrow & Basak, 2011). We focus instead on the growing literature exploring how training on one or more cognitive tasks generalizes, or "transfers," to performance on other cognitive tasks and to daily life. More specifically, we focus on the sorts of claims made by the companies promoting brain-training products and the evidence used to support those claims.

Measures of cognitive performance, including tests of processing speed, reasoning, intelligence, pattern recognition, and similar constructs, have long been used to predict academic and professional success (Deary, 2012; Kuncel, Hezlett, & Ones, 2004). Cognitive and intellectual abilities show stability over time (Kuncel et al., 2004) but also are shaped by experience (Lövdén, Bäckman, Lindenberger, Schaeffer, & Schmiedek, 2010). The promise of cognitive training is based on the following reasoning: If measures of cognitive ability predict real-world performance and success, and if that success depends on those cognitive abilities, then practicing those abilities should improve outcomes—and ultimately improve people's lives.

The term "brain training" is used mostly by companies marketing cognitive interventions to the public rather than by researchers. Throughout our review, we sometimes use the term "brain training" in place of what researchers might call "cognitive training." By doing so, we do not mean to denigrate the scholarly work of those conducting cognitive interventions. Rather, we use the term "brain training" only because it has entered the public lexicon as synonymous with any cognitive intervention that might help to remedy cognitive shortcomings.

Although evocative, "brain training" is somewhat of a misnomer. Only a small fraction of the published studies of the effects of cognitive interventions have assessed neural functioning directly. Claims about the potential effectiveness of cognitive interventions often are grounded in the growing literature on neural plasticity and on the premise that training core cognitive abilities should influence performance on any daily tasks that rely on such abilities (i.e., those that rely on the same brain mechanisms). How brain training affects neural functioning is largely irrelevant to the practical claims made by brain-training companies about how cognitive training improves real-world outcomes. And, in any case, no cognitive change could occur in the absence of some changes to the brain. In essence, the appeal to neuroplasticity provides a "hard science" veneer for the more straightforward claim that practicing one task can improve performance on another one.

We use the term "brain training" in the same way that it is used by these companies: to refer to practicing core cognitive abilities with the goal of improving performance on *other* cognitive tasks, including those involved in everyday activities (e.g., at school, at work, in sports, at home, while driving). Furthermore, we distinguish cognitive-training interventions from studies of skill acquisition and learning—those that focus on how practicing a skill (e.g., music, math) can improve *that* skill. Instead, we focus on transfer of training from one task or skill to another.

The business of brain training

Research on cognitive training and the mass marketing of brain-training products saw a remarkable convergence in the early 2000s. A number of major brain-training intervention papers were published, including the first outcome paper from the largest cognitive-training intervention conducted to date (the Advanced Cognitive Training for Independent and Vital Elderly [ACTIVE] study; Ball et al., 2002), the seminal paper on action-video-game training (C. S. Green & Bavelier, 2003), and the studies that helped launch companies like Cogmed (e.g., Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002). Although several companies promoted cognitive-training products before 2000, most of those were targeted to fairly narrow segments of the population. For example, Scientific Learning Corporation's Fast ForWord product was marketed primarily to those seeking to help children with delayed reading skills.

Although cognitive-training interventions have a long history, the launch of Nintendo's Brain Age in 2005 marked a change in the commercialization of cognitive training. It was the first product marketed to the public on a massive scale, and it brought the idea of brain training into the mainstream. The commercial market for brain-training software has since grown tremendously. Public interest in brain training has grown apace. Recently, a popular book on the subject, *Smarter: The New Science of Building Brain Power* (Hurley, 2013), was endorsed by several leading research psychologists. It described brain-training proponents as "pioneers" and contrasted them with "defenders of the faith," implying that critics of brain-training research are motivated by unreflective faith or temperamental cynicism. "Those arch-skeptics have pretty well lost the argument," said Hurley (Kaufman, 2014).

Most brain-training companies are privately held, so they do not disclose their sales. Even public companies like Nintendo rarely break out the net revenue figures for specific products, so it is not possible to report precise values for the amount of money that consumers spend on brain training even if it is possible to estimate the number of copies sold of certain products. The market research firm SharpBrains (sharpbrains.com) attempts to track all of the significant companies in this industry and publishes market reports. Although SharpBrains does not provide much detail on its methodology and does not provide ranges or confidence intervals around its estimates, it does appear to provide the most comprehensive assessment of the state of the brain-training market.

In their January 2013 report and January 2015 update (SharpBrains, 2013, 2015), SharpBrains assessed the state of the "digital brain health market," including both products for assessing brain function and products for enhancing brain function via training. (SharpBrains also tracks companies that sell hardware-based products, such as products using EEG to assess brain function or transcranial direct current stimulation to enhance brain function.) The firm estimated that the total digital-brain-health market had sales of $210 million in 2005, $600 million in 2009, and $1.3 billion in 2013. In 2013, software accounted for an estimated 55% of the total market ($715 million). In its 2015 update, SharpBrains predicted a total of $6.15 billion in yearly sales by the year 2020. Assuming that the proportion of the market represented by software remains constant, cognitive/brain assessment and training software is predicted to have yearly sales of $3.38 billion by 2020.

The market for brain-training products can be subdivided according to who buys and uses the products. SharpBrains (2015) estimated that 45% of purchases are made by consumers (for themselves personally or for members of their family), with the rest made by employers, schools, or health providers. Of end users, 50% are estimated to be age 50 or over, 30% between 18 and 50, and 20% younger than age 18. Combining these separate estimates from the 2013 SharpBrains report, adult consumers directly spent about $322 million on digital brain-health software products, and if these proportions remain constant, direct consumer spending on such products will reach $1.52 billion in 2020.

Our interest in this review is on products specifically for training, and SharpBrains does not provide separate estimates for the size of the assessment and training markets. However, their January 2015 update listed approximately three times as many "key companies" offering training products for consumers as companies offering assessment products. In a personal communication, the president of SharpBrains (and primary author of its reports) estimated that 60% to 75% of the overall market is for training applications, versus 25% to 40% for assessment (A. Fernandez, personal communication to C. F. Chabris, August 21, 2015).

We used the SharpBrains listing of key companies to constrain our analysis of the evidence for brain training. Specifically, we examined all of the consumer-oriented companies in its listing that market brain-training products, examining their websites for the evidence cited in support of their marketing claims.

The marketing of brain-training products

In the brain-training industry, advertising and marketing efforts often identify or imply benefits that people will receive by using a brain-training product, and people appear ready to accept such claims. In 2014, the AARP released a survey of 1,200 consumers aged 18 and over (David & Gelfeld, 2014). Of those surveyed, 52% were aware of brain training, and of that subset, over 50% agreed that "brain training is exercises or activities that" do each of the following things: "improve memory," "sharpen intellectual skills," "help improve my attention span," "help me think faster," "prevent memory loss," and "increase IQ." Whether advertising of brain-training products led to these beliefs or whether marketing of those products simply taps preexisting beliefs, those who hold such beliefs are likely to believe the claims made by brain-training companies.

As a rule, brain-training companies promote the efficacy of their products for a very wide range of conditions and outcomes, from specific genetic, neurological, and mental diagnoses (e.g., Turner syndrome, age-related cognitive impairment, schizophrenia), to sports performance, general cognitive ability, everyday memory for names and locations, and driving ability. It should perhaps be surprising—and raise doubts regarding the plausibility of the claims—that a single intervention could have such diverse and far-reaching benefits.

In the United States, where the majority of brain-training companies operate, legal regulation of marketing allows for some "puffery"—promotional claims that are subjective *and* unlikely to be confused for objective claims. For example, when a Coca-Cola advertisement shows polar bears enjoying a bottle of Coke, no reasonable person will conclude that Coca-Cola is actually claiming that polar bears consume and like its products (Boudreaux, 1995). In the brain-training industry, advertising statements that seem imprecise by academic or scientific standards could fall within the boundaries of acceptable puffery. However, claims about health benefits or statements drawing on scientific evidence might be subject to other bodies of law and to regulation—see, for example, the FTC's legal action against Lumos Labs regarding the marketing of Lumosity (FTC, 2016a, 2016b, 2016c), in which the parties agreed in a settlement that randomized, controlled, blinded trials would be the standard of evidence for health- and performance-related efficacy claims.

We do not attempt to distinguish between puffery and illegal, deceptive advertising. We also will not attempt to evaluate each advertising claim against the scientific evidence. Instead, we describe how brain-training products are generally advertised, and we evaluate the evidence for the efficacy of brain training. Only when a marketing claim clearly misrepresents a scientific finding will we draw attention to the discrepancy. We leave it to readers to evaluate the general extent of agreement or disagreement between more typical claims and the evidence.

Because of the sheer size of the industry, a comprehensive analysis of the marketing of all brain-training products is beyond the scope of this review. The SharpBrains (2013) market report included a survey of over 3,000 "decision makers and early adopters" who subscribed to an electronic newsletter about the "brain fitness" industry. Respondents to the survey were asked to "name the main product you purchased." The makers of the most frequently named products were Lumos Labs, Posit Science, Nintendo, and Cogmed. For the purpose of illustrating how the industry as a whole markets its products, we will focus on these four well-known companies.

Nintendo. Nintendo's Brain Age was the first product to bring the concept of computerized brain training to a mass market. Its long-used slogan, "Train Your Brain in Minutes a Day!", appeals to the idea that cognitive improvement requires little effort. In the roughly 10 years since it launched, Brain Age and its sequel, Brain Age 2, have sold a combined 34 million copies, making them the fourth- and fifth-best-selling software packages for Nintendo's handheld 3DS System (Nintendo Corp., 2015). In an advertisement that was on Nintendo's website from 2009 through 2012 (and possibly longer), the voiceover narration said, "By completing a few challenging exercises and puzzles, you can help keep your mind sharp" (Japancommercials4U2, 2009, 0:18; previously hosted at "Brain Age on TV," 2012). Nintendo has also used the marketing technique of celebrity endorsement: Actress Nicole Kidman said of Brain Age 2, "I've quickly found that training my brain is a great way to keep my mind feeling young" (Burman, 2007).

The website for Nintendo's more recent 3DS title, Brain Age: Concentration Training, describes its training as "a new training method . . . based on recent brain science . . . an efficient manner of training that always pushes the capabilities of working memory to the limit" ("Brain Age: Concentration Training," 2012). The same page includes a box describing a structural MRI result, with the claim that after 2 months, "working memory training had increased the volume of the cerebral cortex (in mainly the prefrontal cortex)." Next to this claim is a brain image with colored blobs labeled "increased cerebral cortex volume due to training," and below that is the statement "It is said that the cerebral cortex's volume is involved in intelligence. This game contains many kinds of 'devilish training' based on the activities used in that experiment." The website provides no citation to the study or source for the image, but it might be referring to Takeuchi et al. (2011).

On a page accessed by a link labeled "See how it works," Nintendo states that its previous games, Brain Age and Brain Age 2, "prevent decreased brain function by stimulating blood flow in the brain" ("Brain Age: Devilish Training," 2012). By comparison, the new Brain Age: Concentration Training "helps to improve brain function." More specifically, the text says that "Devilish Training can improve your ability to make predictions and decisions, raising your efficiency at work or study, sports, and more." A pair of cartoons asks whether you are "constantly checking e-mail" or "fidgeting without your cell phone," and text below says that "by improving concentration, you can focus more on what is right in front of you."

These advertising claims do not promise measurable, quantifiable benefits, but they do imply significant qualitative benefits (e.g., becoming less distractible in daily life). Although the claims are hedged slightly with phrases like "can improve" or "it is said," they are made more persuasive by appealing to concepts from cognitive psychology and neuroscience (working memory, prefrontal cortex volume, intelligence, "recent brain science"); decorating non-neuroscientific claims with irrelevant "neurobabble" may make the claims more persuasive (e.g., Weisberg, Keil, Goodstein, Rawson, & Gray, 2008), and adding colorful brain images can inflate belief in claims related to neuroscience (e.g., McCabe & Castel, 2008; but see C. J. Hook & Farah, 2013; Michael, Newman, Vuorre, Cumming, & Garry, 2013). Brain structure and function are shaped by experience (e.g., Lövdén et al., 2010; Taya, Sun, Babiloni, & Beszerianos, 2015), but the existence of plasticity does not imply that any particular experiences will produce brain changes that benefit health, learning, or adaptability. The marketing also touts specific mechanisms whereby brain training may improve real-world cognition (e.g., the training will increase the volume of parts of the brain or increase blood flow, or improving working memory will improve decision-making ability) and claims research backing for its claims of efficacy (without citing a study).

Lumos Labs. Perhaps the most pervasively and widely marketed brain-training product is Lumosity, from Lumos Labs. Lumosity advertisements have appeared extensively on television, radio, and the Internet for several years. According to the FTC (2016b), Lumosity's TV ads appeared on at least "44 broadcast and cable networks," and it was also advertised on National Public Radio, Pandora, Spotify, Sirius XM, and via "hundreds of keywords" on Google AdWords. The AARP survey (David & Gelfeld, 2015) found that 51% of consumers were aware of Lumosity. One early commercial featured a man who encountered an old friend on the street but couldn't remember his name, implying that Lumosity training will enhance everyday memory. Internet advertisements for Lumosity regularly appear on web pages for related topics, such as memory and attention—even, ironically, on pages that are critical of brain-training claims. Recently, Lumos Labs partnered with Disney Pixar, and characters from Pixar's animated film *Inside Out* appeared in an ad for Lumosity (Best Funny Commercial, 2015). Such entertainment industry cross-marketing reflects the broad market penetration of brain-training products. Lumosity is large enough to afford high-profile promotions, and public acceptance of Lumosity is such that Pixar is willing to associate its brand with the product.

The voiceover for Lumosity's Pixar ad stated, in full:

> Disney Pixar and Lumosity know that you exercise your body, but what about your mind? It's time to stimulate your mind, with a brand-new challenge. Try Lumosity. You'll get a personalized brain-training program, with games designed by neuroscientists to challenge memory and attention, and keep your mind active by giving its different parts a fun, balanced workout. Visit Lumosity.com to discover what your brain can do. And see Disney Pixar's *Inside Out*, in 3-D, June 19th.

Like Nintendo, Lumos Labs emphasizes the supposed neuroscientific basis of its products as well as their potential benefits for memory and attention. In its recent complaint against Lumos Labs, the FTC noted at least two implicit claims of improvement in brain-imaging outcomes in uncontrolled intervention studies (FTC, 2016b, pp. 9–10). The company's advertisements typically rely on the consumer to supply a "brain as a muscle" metaphor and to infer from general statements (e.g., "stimulate your mind"; "keep your mind active"; "fun, balanced workout") that mental exercise via Lumosity will improve mental function.

On a version of its website available prior to the FTC ruling, Lumos Labs implied more strongly that Lumosity training enhances cognition: "Lumosity exercises are engineered to train a variety of core cognitive functions" ("Welcome! Let's Build Your Personalized Training Program," 2013). Note that the company did not claim that the product would improve everyday cognitive performance, instead relying on the consumer to infer what the verb "train" and the adjective "core" meant in this context. People likely believe that training will lead to improvements in areas broader than just the training task and that "core functions" are those that are the most important and widely useful.

Lumos Labs' website also specified a scientific basis for the product: "Lumosity's groundbreaking program is based on research in the field of neuroplasticity" ("About Lumosity," 2010). At the time we wrote this report, the Lumosity website explained that the product's games are based on well-established tasks in cognitive psychology. It stated, for example, that "for decades, researchers have created tasks that *measure* cognitive abilities" [emphasis added]. The go/no-go task "helps psychologists *evaluate* impulse control," and the Thurstone Punched Holes task "is developed as a cognitive *test*" [emphases added] (see the "We transform science into delightful games" section of "Cutting Edge Neuroscience Personalized for You," 2015). The site does not explicitly claim that the original research tasks trained or improved cognitive abilities. The connection between Lumosity games and the cognitive tasks is not entirely clear in some advertising; for example, the Lumosity game Speed Pack is said to "use similar principles" to the Thurstone task. The academic pedigree of the cognitive tasks as measures appears to be used to imply that a history of high-quality research supports the use of Lumosity games as training tools—an entirely different application.

Posit Science. Posit Science is classified by SharpBrains (2015) as operating in the "consumer" space of the digital-brain-health market. Compared to Nintendo and Lumos Labs, Posit Science engages in less marketing aimed directly at potential users of its products. Instead, it partners with third-party organizations that already have large memberships and/or lengthy relationships with potential users, and these partner organizations advertise and/or deliver the brain-training product.

For example, in 2009, the American Automobile Association (AAA) Foundation for Traffic Safety began providing Posit Science's Drivesharp product at a discounted price to its members (Fernandez, 2009). The United Services Automobile Association (USAA) insurance company offers the online program free to its members, a previous version of its website marketing the program by claiming that "in as little as 10 minutes a day, the Drivesharp program may strengthen your brain's ability to focus on and process what you see while driving. This may enable you to react more quickly to unexpected situations" ("Drivesharp Online Training," 2015). Philips Lifeline, a medical alert service, provided a discounted rate for Posit Science's BrainHQ product to its clients because "large-scale medical trials show that it results in 131% faster processing speed, 50% fewer at-fault crashes, [and] 10 years improvement in memory performance" ("Philips Brain Fitness Powered by BrainHQ," 2015). The AARP, an organization with over 37 million members, now offers a "Staying Sharp" membership that includes a subscription to BrainHQ. Perhaps unsurprisingly, in the AARP's own 2014 survey, 70% of consumers said they would consider using BrainHQ in the future (David & Gelfeld, 2014).

Unlike Lumos Labs and Nintendo, Posit Science provides detailed scientific background material and extensive citations to the scientific literature. However, its marketing uses similar general concepts ("neuroscience," "neuroplasticity,"), draws on appeals to the scientific credentials of its founders, and make claims of the efficacy of its products for improving everyday functioning.

Cogmed. Cogmed is classified by SharpBrains (2015) as operating in the "professional" space. Like Posit Science, Cogmed generally does not advertise directly to the end user. Instead, its working-memory software is distributed to users via health practitioners and schools, where personnel are trained by Cogmed and supervised by "coaches." The company (now owned by Pearson) provides on its website a downloadable report detailing its "claims and evidence," with citations to dozens of publications. Generally, it describes the software as "a computer-based solution for attention problems caused by poor working memory" ("Cogmed Solutions Help," 2015). Thus, it frames the product as being intended to remedy deficits, not to improve already-normal functioning. However, in a paragraph with the heading "Who is Cogmed For?" the company includes in this group any people who "find they're not doing as well as they could, academically or professionally, given their intelligence and their efforts" ("Frequently Asked Questions," 2015). Furthermore, like that of Nintendo and Lumos, Cogmed's marketing refers to "neuroscience," "neuroplasticity," "leading cognitive neuroscientists," and "working memory." Just below a heading on its website stating that "Cogmed does not make extravagant claims," Cogmed claims that "we can confidently state that approximately 80% of individuals with attention and working memory problems who complete Cogmed training will experience meaningful benefits" ("How Is Cogmed Different From Other Cognitive Training Programs?", 2015). Elsewhere, the website states that "the concept of neuroplasticity, the idea that the brain can reorganize itself and change, is what allows Cogmed to effectively change the way the brain functions to perform at its maximum capacity" ("How Is Cogmed Different," 2015). Much like that of Posit Science, Cogmed's website provides detailed scientific background materials and citations to the scientific literature while also making claims about the effectiveness of its products.

Learning, transfer, and the logic of brain training

Most brain-training claims follow from the assumption that practice yields improvements that go beyond the practiced tasks: Playing a video game will enhance basic measures of attention, training on a working-memory task will enhance intelligence, improving speed of processing in a laboratory task will enhance real-world driving performance, and so on. Few doubt that practice improves performance or that the same cognitive mechanisms underlie many skills, or even that practicing one task (e.g., juggling balls) will lead to more efficient learning of closely related skills (e.g., juggling pins). Similarly, our prior experiences prepare us for novel situations: The value of schooling rests on the assumption that learning and applying content knowledge (e.g., about history, science, or literature), skills (e.g., reading), and abstract principles (e.g., algebra) will generalize to nonschool contexts. Even if learners forget the particulars of English authors or benzene rings, the habits of mental engagement may generalize to other settings and materials. The core controversy in the debate about brain training is not about the benefits of practice or about the human potential to develop "academically, socially, and professionally." Rather, it is about whether and when practicing one task will improve performance on *untrained* tasks. The marketing claims of brain-training companies—that practicing their tasks will yield widespread improvements in real-world cognitive performance—are provocative because they diverge from the broader scientific literature on transfer of training.

A brief history of theories of transfer. Most accounts of transfer of training trace their roots to two ideas, each more than a century old: (a) formal discipline theory and (b) transfer by identical elements. Both are based on the idea that transfer depends on a similarity between the content (e.g., knowledge, skill) learned initially and its later application.

According to formal discipline theory, the mind consists of capacities (e.g., concentration, reasoning ability, memory) that can be improved through exercise. This idea traces its roots to Plato's *Republic*. It has been the basis of education systems from classical training in the Middle Ages to the Boston Latin School in the 17th century, and it is still invoked as a rationale for educational requirements today (e.g., freshman algebra).

William James was among the first to explore the theory of formal discipline experimentally, using himself as a subject (James, 1890, pp. 666–668). If memory is a general capacity, he reasoned, then practice memorizing one poem should improve the speed with which he could memorize another poem. He recorded a baseline rate of 50 seconds per line to memorize 158 lines from Victor Hugo's *Satyr*. After many days spent memorizing Milton's *Paradise Lost*, he observed no savings in memorizing a different set of 158 lines from *Satyr* (at 57 seconds per line). Admitting that he was tired from other work for his second bout of memorizing, he recruited four students to repeat the experiment, two of whom showed savings and two of whom did not. Given such weak evidence, James held to the view that "one's native retentiveness is unchangeable" (p. 663).

Using somewhat more rigorous methods in the lab and in controlled classroom research, Edward Thorndike also tested the formal discipline idea. Students who practiced estimating the area of rectangles improved, but their improvements did not transfer to estimating the areas of other shapes. He concluded that "the mind is so specialized into a multitude of independent capacities that we alter human nature only in small spots" (Thorndike, 1906, p. 246), a view often referred to as "transfer by identical elements," whereby skills acquired during training are tightly coupled to the stimuli, tasks, and responses required during learning:

> Training the mind means the development of thousands of particular independent capacities, the formation of countless particular habits, for the working of any mental capacity depends upon the concrete data with which it works. Improvement of any one mental function or activity will improve others only in so far as they possess elements [in] common. . . The most common and surest source of general improvement of a capacity is to train it in many particular connections. (p. 248)

During the early 20th century, these two theories inspired heated debate and intense empirical study (Angell, Pillsbury, & Judd, 1908; Bagley, 1909; Bennett, 1907; Hewins, 1916). Although Thorndike's arguments won the day, with most contemporary reviews rejecting formal discipline, the empirical evidence provided more support for it than is often recognized today. For example, high school biology students who practiced their observational skills (e.g., by writing descriptions of flowers and leaves under a time limit) later wrote more detailed descriptions of novel biological specimens and showed greater improvements in non-biological tasks (i.e., drawing a nonsense figure from memory) than did control participants who just studied the textbook (Hewins, 1916).

Although formal discipline and identical elements are often presented as conflicting theories of generalization from learning, they are not so easily distinguished empirically because neither theory provides operational definitions that distinguish capacities from elements. For example, Thorndike (1906) argued that learning Shakespearean sonnets should improve one's ability to remember Bible verses because both rely on similar elements of memory (p. 241). Yet savings in memorizing new poetry was the very phenomenon that James discounted as a failure of formal discipline theory (also cf. Bagley, 1909). Similarly, Hewins's findings of transfer might be attributed to the training of particular cognitive elements (e.g., feature selection, verbal coding, fluency) common to the learning and transfer tasks. In broad strokes, then, the difference between theories of formal discipline (i.e., training capacities or abilities) and identical elements (i.e., training cognitive components) may be a matter of scale. The field still lacks a comprehensive theory of transfer, so the question of when and where we apply what we learn is largely unresolved (Barnett & Ceci, 2002). Caricatures of formal discipline as the idea that memorizing poetry will have broad effects on intellect can be dismissed (Roediger, 2013), as can caricatures of identical elements as the idea that all learning must be highly grounded in everyday experience (J. Anderson, Reder, & Simon, 1996).

In contemporary literature, the language of transfer largely draws on Thorndike's idea about elements (cf. Barnett & Ceci, 2002). Transfer tasks that share many elements with the practiced task are said to illustrate near transfer, whereas tasks that share fewer elements are said to illustrate far transfer. We can make an orthogonal distinction between (a) horizontal transfer, in which the elements acquired at learning are similar in complexity to those of the transfer task (e.g., learning how to use one word processor will facilitate learning how to use another one, without much difference in which was learned first), and (b) vertical transfer, in which the elements at learning are somewhat simpler than those at transfer, making the order of learning important (e.g., readers must learn to recognize letters and map them onto symbols before they can read words and sentences; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Modern theories define elements in terms of the logic of production rules (Singley & Anderson, 1989; Taatgen, 2013), whereby a skill can be modeled as a set of conditional (if-then) statements defining how particular conditions (e.g., stimuli, mental states) lead to particular outcomes (e.g., behaviors, changes in mental states). The advantage of defining elements in this way is that it enables simulations that can predict when transfer will and will not occur. For example, based on simulation results, Taatgen (2013) has argued that practice with task switching can transfer to tasks such as working-memory and Stroop tasks.

Formal discipline contributes to contemporary conceptualizations of learning as well, particularly in research on reasoning, causal inference, and logic (Nisbett, Fong, Lehman, & Cheng, 1987). For example, learning abstract principles in logic and statistical reasoning transfers to the solution of novel problems, especially when the abstract principles are coupled with intuitive, worked examples (Lehman & Nisbett, 1990). Similarly, 2 years of graduate study in a science that requires probabilistic reasoning (e.g., psychology, medicine) resulted in bigger improvements in conditional (if-then) and biconditional (if-and-only-if) reasoning than graduate study in a science that relies more on deterministic principles (i.e., chemistry; Lehman, Lempert, & Nisbett, 1988).

Modern theories of transfer recognize the importance of context and the potential for interactions between the *content* of what is learned and the *contexts* in which learning and transfer occur. Barnett and Ceci (2002) proposed a taxonomy to characterize how content and context can interact to determine transfer. For example, learned skills can vary in their specificity, from highly routinized procedures to more abstract principles or heuristics. The context of learning and transfer can vary along many dimensions, including knowledge domain, physical location, intended purposes (e.g., academic, work-related), and whether demonstration of the skill requires only the individual or involves other people. In this view, transfer depends on the content of learning, the similarity of the contexts in which that learning is applied, and the interaction between the content and context. Highly specific content, such as a routine procedure, should show less transfer than broader content, such as a strategy. At the same time, success in applying principles or strategies is likely to depend on the context.

What we know about learning and transfer. Practice generally improves performance, but only for the practiced task or nearly identical ones; practice generally does not enhance other skills, even related ones (see Noack, Lövdén, Schmiedek, & Linderberger, 2009; Stine-Morrow & Basak, 2011, for reviews). For example, performance on inductive and spatial reasoning tasks is moderately correlated (and both are examples of "fluid" abilities), but training in inductive reasoning does not enhance spatial reasoning, nor does training in spatial reasoning enhance inductive reasoning (Blieszner, Willis, & Baltes, 1981; Plemons, Willis, & Baltes, 1978; see also Ball et al., 2002).

The lack of transfer from one content domain to another is also seen in the development of expertise. A person who practices memorizing digits will increase his or her digit span but will be unlikely to show a benefit for color span or letter span (Ericsson & Kintsch, 1995). Chess grandmasters can recall a mid-game chess position with remarkable accuracy after viewing it for only a few seconds (e.g., de Groot, 1946/1978), but they show little advantage when remembering other types of materials. In fact, their recall is barely better than that of an amateur for chess pieces that are positioned randomly on a board (Chase & Simon, 1973; Gobet & Simon, 1996). As a general observation, empirical examples of near transfer, in terms of both content and context, are more prevalent than those of far transfer (Barnett & Ceci, 2002; Postman, 1972; Singley & Anderson, 1989).

This specificity of effects tends to hold true for studies of neural plasticity as well; practice effects on neural growth tend to be specific to the neural substrates of the behaviors involved in the practice (Draganski et al., 2006; Draganski & May, 2008; Maguire et al., 2000; Maguire et al., 2003). For example, violinists show selective neural growth in the right motor cortex, corresponding to the use of their left hand to finger the strings (Herholz & Zatorre, 2012).

Perhaps the strongest evidence for transfer from learning to meaningful outcomes comes from engagement in formal education (see Stine-Morrow & Payne, 2015, for a review). Educational attainment, intelligence, socioeconomic status, and income are intercorrelated, so determining the causal effects of schooling on cognition is complicated. However, evidence from "natural experiments" (e.g., accidents of geography or history that impact educational opportunities, largely without regard for talent or individual resources) and large-scale longitudinal studies provide some support for the causal effect of education on intellectual development. For example, kindergarteners whose birthdays fall just after the cutoff date to begin school show better scores on tasks of verbal working memory and inhibitory control than do prekindergarteners whose birthdays fall just a few days earlier (Burrage et al., 2008). And a longitudinal study initiated in the 1930s that involved testing almost all of the 11-year-olds in Scotland (the Lothian Birth Cohort study) found that education level predicted intelligence later in life, even after controlling for intelligence and socioeconomic status in childhood (Ritchie, Bates, Der, Starr, & Deary, 2013). These changes appeared to result from improvements in particular cognitive abilities rather than intelligence as a whole (Ritchie, Bates, & Deary, 2015).

Not all evidence supports an education-intelligence link, however. Students in Massachusetts who were assigned to a charter school based on a lottery showed greater increases in scores on standardized assessments of math and language arts than did those assigned to traditional public schools. However, those improvements did not transfer to measures of working memory or nonverbal intelligence (A. S. Finn et al., 2014). In other words, the students learned what they were taught, but those skills did not transfer to more basic measures of cognitive capacity. Of course, more research is needed to determine how and when education yields generalizable skills in addition to enhanced knowledge of the studied content (Stine-Morrow, Hussey, & Ng, 2015).

Nevertheless, we know of no evidence for broad-based improvement in cognition, academic achievement, professional performance, and/or social competencies that derives from decontextualized practice of cognitive skills devoid of domain-specific content. Rather, the development of such capacities appears to require sustained investment in relatively complex environments that afford opportunities for consistent practice and engagement with domain-related challenges (Ericsson, 2006; Ericsson & Charness, 1994; Ericsson & Kintsch, 1995; Ericsson, Krampe, & Tesch-Römer, 1993; Grossmann et al., 2012; Rohwedder & Willis, 2010; Schooler & Mulatu, 2001; Schooler, Mulatu, & Oates, 1999, 2004; Shimamura, Berry, Mangels, Rusting, & Jurica, 1995; Simonton, 1988, 1990, 2000; Staudinger, 1996; Staudinger & Baltes, 1996; Staudinger, Smith, & Baltes, 1992; Stern, 2009), a factor that brain-training programs largely ignore. The development of these capacities often also relies on reasoning, judgment, and decision-making, and generalization of even the simplest forms of reasoning to new situations requires multiple exposures to the content and practice in those new contexts (Gick & Holyoak, 1980, 1983; Wertheimer, 1945/1959). Thus, the gap between the skills trained by brain-training software and their target applications is often large.

Given the general lack of evidence for the development of cognitive capacities based on short-term experiences of the sort that can be studied in the lab, the current research literature on learning and transfer has largely focused on the principles of learning and motivation that engender effective acquisition of new information and new skill sets, and on the ways in which instruction can be organized to promote the application of that knowledge (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Healy & Bourne, 2012). In other words, learning and transfer are themselves acquired skills (Brown, Roediger, & McDaniel, 2014). Interestingly, the conditions that maximize efficient learning often contrast with those that lead to durability and broader application of that learning. Relatively consistent practice, immediate testing, and frequent and consistent feedback all can speed initial learning but impair retention and transfer relative to learning under more varied conditions with delayed testing at variable intervals (Schmidt & Bjork, 1992). Practiced routines can become exceptionally efficient, but that efficiency can come at the cost of generality. In contrast, effortful strategies that involve self-explanation and self-generated retrieval of learned material, especially with spacing, can enhance the durability of memory and learning (Dunlosky et al., 2013).

Summary and implications. Brain-training programs typically train performance on relatively simple skills in a limited range of contexts (typically on a home computer and with little involvement of substantive content or knowledge), but their marketing materials imply generalization to a wide range of skills in varied contexts with varied content. For example, the website for Posit Science's BrainHQ software cites testimonials from users claiming that "BrainHQ has done everything from improving their bowling game, to enabling them to get a job, to reviving their creativity, to making them feel more confident about their future" ("Brain Training That Works," 2016). Brain training seems to depend on the logic of vertical transfer; the programs purportedly exercise component skills that can be applied in substantive content domains and in different contexts (far transfer), a claim that runs counter to evidence for the narrowness of transfer across content and contexts. Moreover, this logic is predicated on an incomplete model of the vertical path between the trained skills and the outcome measures.

Modeling this path is not trivial. Although conceptualizations of transfer have become more nuanced, a fundamental problem remains: defining similarity across content and context (Barnett & Ceci, 2002; Noack et al., 2009). Consequently, when evaluating whether transfer is "near" or "far," we are left largely with relative judgments. Although most brain-training companies describe a rationale for their choice of training methods, they rarely discuss the mechanisms of transfer explicitly, other than by making a general appeal to the concept of neuroplasticity and the notion that "the brain is like a muscle." Few companies discuss the importance of motivation or context in the likelihood of transfer or consider how their particular training context might enhance or limit transfer to other contexts. In our view, if speed-of-processing training on a simple laboratory task improves speed of processing in another laboratory task, that would constitute relatively near (and horizontal) transfer; improved speed of decision-making in driving, on the other hand, would suggest relatively far (and vertical) transfer, because it involves training a component skill and transferring to a more

Table 1. Brain-Training Companies Listed by SharpBrains, Grouped by the Evidence Provided for the Effectiveness of Their Products

Brain-training companies citing multiple publications reporting tests of the effectiveness of a marketed brain-training product	CogMed CogniFit Lumos Labs (Lumosity) Posit Science (BrainHQ) Scientific Learning Corporation (Fast ForWord)
Brain-training companies citing some intervention research, but not necessarily tests of the effectiveness of a marketed brain-training product	Advanced Brain Technology Akili Interactive Labs ACE Applied Cognitive Engineering Dakim Brain Fitness Learning Enhancement Corporation Neuronix Scientific Brain Training (HAPPYneuron)
Brain-training companies citing no peer-reviewed evidence from intervention studies	Brain Center America [possibly defunct] Braingle BrainMetrix Brain Resource (myBrainSolutions) C8 Sciences Cognisens Focus Education Games for the Brain Houghton Mifflin Harcourt Earobics Nintendo Brain Age Peak Vivity Labs (FitBrains)
Companies conducting training or promoting products that fall outside the scope of this review	Blue Marble e-hub Happify Mindset Works Ultrasis Vigorous Mind

complex one in a different context. If working-memory training improved performance on another working-memory task, that would be nearer transfer than improved performance on an intelligence test (greater difference in content). Throughout this review, we will make such relative judgments of the extent of transfer, but readers should keep in mind that, in the absence of a complete understanding of the nature of these underlying mechanisms and without a full description of the role of context, these judgments are necessarily somewhat subjective.

The Scope of This Review

For this review, we focus exclusively on published, peer-reviewed, scientific journal articles cited by brain-training companies and proponents as support for the scientific credibility of their claims. These include 132 papers identified by the signatories of the open letter in support of brain-training effectiveness ("Studies on Cognitive Training Benefits," 2014). We supplement that list with all of the published articles cited as supportive evidence on the websites of leading brain-training companies identified by SharpBrains. Its review tracks companies that have substantial "market and/or research traction." Of the more than 200 companies in the brain-training industry, the SharpBrains (2013) report and 2015 update identified 30 training companies as those with market or research traction. Of those 30 training companies, six promote interventions (e.g., physical fitness) or outcomes (e.g., mental health, well-being, or happiness) that fall outside of our scope. Although almost all of the remaining 24 companies tout their scientific credibility, with some even claiming "proven" effectiveness, the websites of half of these companies cite no peer-reviewed scientific evidence for the effectiveness of their interventions (see Table 1). Those lacking citations to published research generally appeal to the expertise of their founders and scientific advisors, post testimonials from customers or clients, or provide internal reports or assessments of their program's effectiveness.

The 12 companies that did cite scientific publications varied widely in both the quantity and relevance of those citations. Some companies cite basic research on neuroplasticity or on the cognitive tasks that form the basis of their training but cite no papers documenting the effectiveness of their own programs. Such citations to studies on other products provide no direct support; they are akin to a drug company promoting its own product based on another (perhaps related) drug's effectiveness in clinical trials—or based on biological laboratory findings about the mechanisms of action thought to underlie the drug's effects.

A few companies stand out for their efforts to document the scientific credibility of their claims, citing tests of the effectiveness of their own products or programs. These include Cogmed, CogniFit, Posit Science, and Scientific Learning Corporation. Lumos Labs, the maker of Lumosity, also provides an extensive list of citations, but relatively few are to peer-reviewed research. For our review, we included all papers cited on the websites of these companies as of January 2015, even if the cited studies did not test the effectiveness of that company's products.

The citations provided by the Cognitive Training Data website and by the companies themselves presumably represent a collection of those articles that leading companies and cognitive-training researchers believe best support the effectiveness of brain-training interventions and that best justify the claims made by companies. This list is incomplete, especially given the rapid rate of publication of cognitive-training articles, but it arguably provides a fairly comprehensive set of those articles that leading brain-training researchers believe best support the effectiveness of brain-training interventions; the signatories included the founders of several prominent brain-training companies as well as board members and consultants for many brain-training companies.

In addition to the research cited by companies and proponents of brain training, we examined two major areas of cognitive training that have been conducted largely independently of brain-training companies: adaptive *n*-back working-memory training and video-game training. Both of these literatures are cited extensively by brain-training companies—many of which incorporate video-game-like or working-memory tasks into their products—but emerged independent of the development of commercial products.

To ensure that the citations gathered from company websites were not affected by the date we happened to conduct each search, all searches were conducted in 2015. That ensures complete coverage of papers cited by brain-training companies prior to 2015. Our search does include some papers from 2015, depending on when we searched each company website.

Note that by restricting our analysis to the published scientific literature, we have risked neglecting negative results that often are presented at conferences but not published in journals: Any review focusing exclusively on the published literature is more likely to detect reports of successful than unsuccessful studies as a result of biases against the publication of null findings (Ioannidis, 2005). The complete list of articles that we compiled from brain-training companies and Cognitive Training Data is available on the Open Science Framework page for this project (https://goo.gl/N6jY3s).

Best Practices in Intervention Design and Reporting

Before assessing the evidence for brain-training benefits, we must specify the criteria we used to evaluate study quality. To that end, in this section, we discuss the inferences permitted by different study designs, the practices that strengthen or undermine conclusions about an intervention, and broader problems in scientific practice and publishing that affect the quality of published research. The guidelines and best practices we describe are not specific to cognitive-training interventions—they apply to any psychological intervention studies—but we will focus on how well cognitive-intervention research meets these standards.

The gold-standard design

When evaluating the effectiveness of any treatment, drug, or intervention, the gold standard is a double-blind, placebo-controlled, randomized clinical trial. In that design, participants are randomly assigned to a treatment group or an appropriate control group; they do not know whether they have received the treatment or the placebo, and the person conducting the testing does not know whether any particular participant is in the treatment or the control condition. Whenever inferences about a treatment depend on generalizing from a sample of participants to a larger population, this design is the gold standard for inferring causality.

With large enough samples, random assignment to conditions helps to ensure that the experimental and control group are roughly equivalent in most respects—on average, the participants assigned to each condition will be equalized on extraneous factors that could affect the outcome. And, across many such studies, the average effectiveness of an intervention should be unaffected by such uncontrolled factors. Blinding both participants and testers to condition assignment helps eliminate any effects of expectations or systematic biases from inducing performance differences between the intervention condition and the control condition. With random assignment, suitable blinding, and an appropriate control group, differential improvements between the experimental and control group can be attributed to the critical aspect, or "ingredient," of the intervention itself (i.e., the specific way in which the intervention group is treated differently from the control group). When conducted appropriately, this intervention design permits causal claims about the effectiveness of the treatment.

Other designs and their limitations

Other designs might suggest a link between a treatment and an outcome, but they typically do not permit unambiguous causal inferences about the effect of that treatment. This section considers several other types of studies whose results are commonly cited as evidence for the effectiveness of brain training and discusses their limitations.

In the early stages of research in a new field, more limited designs are common. They can provide preliminary suggestions that later are tested using more rigorous standards. The designs range from observational to correlational to experimental, but all have shortcomings that weaken the defensible conclusions that can be drawn from them.

Correlational designs and skilled-versus-novice comparisons. For elderly drivers, performance on a measure of attention and processing speed called the Useful Field of View Test (UFOV) is correlated with crash rates, such that participants with lower UFOV scores are more prone to crashes (Clay et al., 2005). Correlational findings like this one provide an intriguing link between a laboratory measure and real-world performance, but that does not mean that improving performance on the UFOV will yield better driving performance. For example, both driving performance and UFOV performance might be influenced by some other factor (a third variable), such as sensory-perceptual abilities, motor control, resistance to distractibility, general health, or any of the many other factors that change with age. If so, training with the UFOV might have no impact on driving performance—it could improve UFOV performance but not the underlying causal factor linking the UFOV to driving. And, even if a correlation between two measures does result from a causal relationship, improvements on one task need not produce improvements on the other. Imagine an extreme example in which "training" consisted in providing participants with the appropriate response for each UFOV trial in advance. Performance on the UFOV would improve dramatically, but those improvements likely would not transfer to driving performance or to anything else. More generally, the presence of a correlation between a cognitive task and an outcome measure does not guarantee a causal link between the two, and correlational studies should not be taken as evidence that training on one of the measures will improve performance on the other.

A related approach involves comparing skilled performers and novices. For example, experienced video-game players sometimes perform better on measures of cognition and perception (e.g., they have greater attention breadth; C. S. Green & Bavelier, 2003). But such differences between skilled and novice players alone provide no evidence that the difference in skill caused the superior performance of experienced players on other tasks, so they provide no direct evidence for the effectiveness of brain training. In fact, the causality just as plausibly could go in the opposite direction: People might become video gamers precisely because they can respond quickly and focus attention broadly, allowing them to excel at games and motivating them to play more. Because designs based on preexisting groups do not permit random assignment of participants, there might be systematic differences between the groups on variables other than those of interest, and those other variables might drive any differences.

Correlations or group differences are consistent with evidence of brain training: If there were no link between the UFOV and driving performance, there would be no reason to think that training with the UFOV would enhance driving performance. If skilled and novice gamers did not differ in their performance on a cognitive task, there would be no reason to expect game training to improve cognitive performance. But only an appropriately conducted training study can provide evidence for the causal role of that difference.

Longitudinal designs. Another way to look for potential benefits of brain training involves observing large numbers of people and measuring or predicting cognitive outcomes from their typical daily activities. For example, a study of more than 4,000 older adults in Chicago found that those who reported taking part in more cognitively engaging activities at baseline (e.g., reading, playing games, going to museums) showed more preserved cognitive functioning on standard neuropsychological measures over a 5-year period (Wilson et al., 2003). Similarly, greater educational attainment (e.g., years of education completed) predicts smaller declines in cognitive functioning over a period of years (e.g., Albert et al., 1995). Such results are consistent with the idea that engaging in cognitive tasks preserves cognitive functioning, but like other correlational and between-group designs, such longitudinal studies do not control for all confounds that might explain these relationships. Moreover, even if the relationship is causal, the directionality is often unclear; baseline cognitive performance might predict whether or not people engage in cognitively demanding activities (Bosma et al., 2002). Outside a true experiment, observed relationships between cognitive engagement and cognitive functioning, even those tested with relatively large samples, should be interpreted cautiously.

Experimental designs with shortcomings. Even the use of an experimental intervention does not guarantee compelling evidence for the effectiveness of a brain-training regimen.

The need for an appropriate control group. Imagine a brain-training study in which all participants take a pretest of their driving performance and then complete 10 hours of training on a task designed to measure their ability to spread attention to multiple locations (attention breadth). After training, they redo the driving test and perform better than they had at baseline. Can we conclude that attention-breadth training improved their driving performance? No. Without a comparison to an appropriate control condition, we cannot infer that training improved performance. In fact, the attention-breadth training might have had no effect at all, even if driving performance improved dramatically.

An analogy to the effectiveness of a drug might clarify the problem. Imagine you have an ear infection and your doctor prescribes antibiotics. Several days later, your ear infection is cured. Did the antibiotics cause that improvement? Perhaps, but you also might have improved without any treatment. The same is true for cognitive training. Participants might have done better the second time they took the driving test because they had experienced it once before; performance on a task tends to improve with familiarity and practice. Training interventions must control for such practice effects (known as *test-retest improvements*) in order to infer that the training itself mattered.

Many other factors could contribute to improved driving performance in our hypothetical training study, even if attention-breadth training had no effect. For example, participants spent 10 hours in a laboratory, and the social contact with and attention from the experimenter could have contributed to better performance. After engaging in 10 hours of training, participants might be motivated to perform well on the post-training test, yielding an improvement relative to their pre-training test in order to show that their time was well spent. Furthermore, the experimenters might induce an expectation for improvement that itself leads to improvement, a placebo effect. Intervening events could also contribute to changes from pretest to posttest (e.g., seasonal changes; political, societal, or economic events). Without an appropriate baseline control group, improvements following an intervention cannot provide compelling evidence that the intervention helped.

The biggest challenge in any intervention study involves selecting an appropriate control condition as a baseline. An ideal baseline condition should be identical to the treatment condition in all respects except for the critical, "active" ingredient of the treatment. Otherwise, other differences between the groups could account for any differential improvements. Furthermore, the control group should be comparable to the experimental group before treatment or the results of the intervention will be effectively uninterpretable—that is, differential gains after training could just reflect those different starting points (Redick & Webster, 2014).

Unfortunately, many brain-training studies use control conditions that fall short of this standard (Boot, Simons, Stothart, & Stutts, 2013). Weak or poorly matched control conditions might be acceptable when first exploring the possible effectiveness of an intervention, but they limit the strength of the evidence for the power of the intervention.

In a *no-contact* control condition, for example, participants complete the pretest and the posttest but receive no training or other social contact between those tests (a *treatment-as-usual* control is similar except that participants in the control group might have contact with experimenters as part of their normal activities or therapy.) A *waitlist* control group is identical to a no-contact group except that participants are promised the treatment or training in the future. Such *passive* control conditions help to account for test-retest effects (improvements when people complete a task a second time). When participants are randomly assigned to the treatment condition or a passive control condition, the control condition also equates for the passage of time between the pretest and posttest, thereby matching the intervention and control groups for any events that occurred between the pretest and the posttest. However, passive control groups do not account for increases in motivation resulting from the training itself or for the social contact involved in completing the training. Moreover, they do not account for the possibility that doing something extra, even if it was not a brain-training task, might contribute to improved performance. In short, they do little to account for the placebo effect—the possibility that expectations for improvement might actually yield improvements—even if the training itself were ineffective in changing performance.

A more rigorous intervention design incorporates an active control group, one in which participants complete some other activity in place of the training intervention. An active control group makes the experience of participants in the baseline condition more comparable to that of those in the intervention condition, potentially equating the social contact experienced during the training period and reducing motivational differences between the groups. However, just because a control group is active does not mean it is adequate.

Consider, for example, a brain-training study in which the intervention group completes 10 hours of an adaptive working-memory task, one that increases in difficulty as performance improves. Ideally, a control condition for this intervention should place similar demands on the participants but should lack the critical ingredient thought to enhance performance on other outcome measures. If the hypothesized critical ingredient is working memory, then the control condition should include all of the elements of the intervention—the adaptive difficulty, the need for vigilance and effort, and so forth—except for the working-memory component thought to enhance other cognitive measures. Only by equating everything except that critical ingredient is it possible to infer that greater improvements in the intervention group were due to that ingredient. If the control-group treatment was not comparably engaging and challenging, then any improvements could be due to greater engagement, for example, and not to the practice on working memory in particular.

Many control groups are called "active" because the participants do *something* in place of the brain-training intervention. But they are not necessarily active in the critical way. Without appropriate matching of the demands upon and experiences of participants in the training and control groups, strong conclusions about the efficacy of the intervention are unmerited. As an admittedly absurd example, a control condition consisting of 10 hours of flagellation with wet noodles is unquestionably active, and it might control for at least some forms of social contact, but it is a poor control condition for an adaptive working-memory protocol. To our knowledge, no study has used a wet-noodle control group, but many psychology interventions use control groups that are poorly matched to the intervention group in terms of task demands (Boot, Simons, et al., 2013). A common example in the cognitive-intervention literature is the use of an "educational" control in which participants watch a DVD or read related materials rather than using brain-training software. Another example is having participants in the control group solve crossword puzzles. Although such control conditions are technically "active," they generally do not match the demands of an intensive cognitive intervention, and they differ in ways other than just the targeted critical ingredient.

The need to equate expectations. Even with perfectly matched active control groups that differ only in the critical ingredients thought to affect performance, concluding that the ingredient causes differential improvements might still be premature. Interventions in psychology face even bigger interpretive challenges than those in many other disciplines.

Imagine a drug trial in which participants knew whether they had received a sugar pill (placebo) or an experimental drug. Now imagine that the group receiving the actual drug improved at a greater rate. What would you conclude? Awareness of the condition assignment in this study undermines the ability to control for the placebo effect. Without blinding to condition assignment, participants who knew they received the experimental drug would expect greater improvements than would those who knew they had taken a sugar pill.

In a double-blind design, participants do not know whether they have received the drug or the sugar pill, so they have no reason to expect different outcomes. Consequently, any differential improvement can be attributed to the difference in the content of the pills rather than differences in expectations. The purpose of a placebo control group is to equate for expectations, leaving only the difference in the intervention itself as an explanation for differential improvement.

In brain-training interventions (and most psychology interventions), participants are not blind to their condition assignment. If you spend 10 hours training on a working-memory task, you know that you have done so. If you spend 10 hours watching educational DVDs, you know that, too. Figuratively speaking, in brain-training interventions, people know which pill they received, and they likely have expectations for how that pill will or will not help their performance on each outcome measure. They might not know that other participants received different pills, a form of blinding commonly used in well-designed interventions, but they are not blind to the contents of their own pill, or to the fact that they took one.

A crucial consequence of this departure from a true double-blind design is that a classic placebo control group is not possible. Psychology interventions cannot equate expectations by eliminating awareness of (or speculation about) the nature of the intervention. And, perhaps as a result of pervasive positive messages about brain training in the media, people tend to expect brain-training interventions to affect cognitive performance and daily activities, with older adults being more likely than young adults to expect benefits (Rabipour & Davidson, 2015). Consequently, the only way to ensure that expectations are equivalent in the intervention condition and the control condition is to measure those expectations. More precisely, to conclude that brain training—and not a differential placebo effect—enhanced performance on an outcome measure, researchers must demonstrate that the intervention group and control group had comparable expectations for improvement *for that outcome measure* (Boot, Simons, et al., 2013). If expectations for an outcome measure track the pattern of improvements between groups, then it is not possible to isolate the effects of the intervention from the effects of expectations. In such cases, the intervention might well be effective, or it might work only in conjunction with expectations, or it might have no effect at all. The cause of the improvement is indeterminate, and strong claims of a benefit from the intervention itself are not justified.

Although the use of such weaker designs might be justified in the early stages of discovery, it is more likely to yield spurious conclusions about the benefits of an intervention. Even in drug testing, the lack of a suitable placebo control can lead to false conclusions about the efficacy of a drug. Early stages of testing for Food and Drug Administration (FDA) approval of a new drug do not require the rigor of the final testing stage. Yet Phase III testing does require preregistered, large-scale studies with double-blind, placebo-controlled designs. And, according to the FDA (U.S. Food and Drug Administration, 2016), of the "promising" drugs that survive the two first, less-rigorous phases of testing and enter Phase III studies, only 25% to 30% show enough of a benefit relative to an appropriate placebo control that they are allowed to proceed to market. That is, the vast majority of promising drugs show no benefits when more rigorous testing and reporting standards are required. Yet psychology interventions have no equivalent of a Phase III trial. Consequently, they must control for placebo effects in other ways. Few studies in the cognitive-intervention literature have done so adequately.

Best practices in study documentation and reporting

In medicine, all government-funded clinical trials in the United States conducted after the year 2000 have been required to be preregistered on ClinicalTrials.Gov, although these preregistrations vary substantially in their precision and completeness. In a preregistered study, the nature of the experimental intervention, all of the conditions, and all outcome measures are described fully before any data collection begins. The preregistration documentation includes a complete analysis plan, explaining how each outcome variable will be coded and analyzed. Sites like the Open Science Framework (www.openscienceframework.org) and ClinicalTrials.Gov host these plans in a read-only, time-stamped format to provide verification of the preregistration. Some journals, including *Psychological Science*, have begun marking preregistered studies with a badge to indicate that the design and analyses were specified before data collection (Eich, 2014).

Studies lacking preregistration leave open the possibility of flexibility in reporting that can undermine the interpretability of a finding. Imagine a well-designed intervention that finds a bigger improvement in the brain-training condition than in an active control condition for three of its 20 outcome measures. With 20 significance tests and an alpha level of .05, we should expect one significant result, on average, even if none of the outcome measures actually differed between the training group and the control group in the tested population. That is, we expect a false positive rate of 5%. We could correct for multiple comparisons by adjusting the alpha level and treating as statistically significant only those results with p values less than the corrected alpha level (i.e., rather than using $p < .05$ as the criterion for statistical significance, dividing that number by 20 to keep the experiment-wide false-positive rate at 5% even with multiple comparisons—many such correction techniques are available). But suppose the authors write a paper that reports only the three results that were significant and makes no mention of the 17 that were not. Readers would have no way to evaluate the results of the intervention because they could not appropriately correct for multiple comparisons.

Such underreporting can have damaging effects on the reliability of published evidence. A recent analysis of large clinical trials supported by the National Heart, Lung, and Blood Institute (NHLBI) between 1970 and 2012 found that 57% of studies from before 2000 reported positive effects (Kaplan & Irvin, 2015). Since 2000, when federal guidelines added the preregistration requirement, only 8% have reported success. Of course, the lack of a statistically significant result does not necessarily prove the absence of an effect, but this remarkable change raises the possibility that many of the previous positive results resulted in part from the use of weaker and more flexible design and reporting standards (as well as from the failure to publish negative results).

In contrast to medicine, in which all recent trials are preregistered, brain training has seen few preregistered interventions. Moreover, most registrations that do exist lack specificity about the design and analysis plan, and some studies were registered only after data collection had been started or even completed. In the absence of preregistration, the results of published brain-training studies should be interpreted with caution. Even if most published studies show large benefits, the results could be misleading in the same way that the pre-2000 NHLBI clinical trials appear to have been.

Another problem resulting from the absence of preregistration is the challenge of determining which papers report on the same intervention. Large-scale training studies often report outcome measures separately across a number of publications; two different papers might each report one or more outcome measures. Imagine the extreme case of an intervention with 100 distinct outcome measures, five of which are statistically significant at $p < .05$. The results in this case would provide no evidence for the effectiveness of the intervention—the number of significant results is no greater than we would expect by chance. Still, if a researcher were to publish each of the significant outcome measures in a separate manuscript without mentioning the other 99 measures, readers would be left with the impression that the intervention was an effective way to improve performance on those five measures, and an extremely effective intervention in general, with five successful trials and no failures.

Whenever a paper lacks preregistration and does not explicitly identify all of the outcome measures, the results of any statistical significance tests are difficult to interpret. Moreover, unless the authors make it clear that separate papers are based on the same intervention, readers might mistakenly believe that the results came from separate, independent interventions.

The combination of large-scale projects, journal page limits, and other considerations often necessitates publishing different aspects of a study in separate journal articles. For example, a large intervention might include both behavioral and neuroimaging outcome measures, and reporting those outcomes separately might be the best way to reach the relevant target audiences. However, unless such papers explicitly cite each other and make clear that all of the measures were from the same intervention, they mislead readers by giving the false impression that the papers report independent studies or interventions. A failure to alert readers to this non-independence of papers is a serious problem because it undermines the possibility of evaluating the strength of evidence across a field (via meta-analysis or other techniques). Whenever separate papers report the results of a single intervention, they must explicitly note that overlap, and each paper should identify all measures from the intervention, even those reported elsewhere.

Given the scale and expense of brain-training studies, measuring the effect of an intervention on multiple cognitive constructs makes sense. Readers should be skeptical whenever a paper reports only one or two outcome measures following a costly intervention—such papers likely are not reporting all of the outcome measures, and the results are more likely to be spurious. Moreover, if other papers by the same team report different outcome measures from what appears to be the same intervention, the results and conclusions of each paper must be viewed with skepticism. Such a pattern of publication implies that other outcome measures—those that did not "work"—might never be reported, thereby muddying the interpretation of all of the published results.

These sorts of underreporting are just the simplest ways in which brain-training papers might give the impression of effects that are more robust than they actually are. In our hypothetical examples, we assumed that each outcome could be measured in a single way, with one unambiguous statistic. What if each outcome could be measured in multiple ways? For example, attention breadth could be measured using either accuracy or response latency; measures of central tendency can be calculated with or without removing outliers or with or without data transformations. Any such analysis decision might be entirely justifiable and rational, but without preregistration of the analysis plan, it is impossible for a reader to know whether such decisions were planned or were devised after inspecting the data. Analysis decisions made after viewing the data undermine the interpretation of the reported statistical significance tests. Even with the best of intentions, the multiple paths to a final paper introduce investigator degrees of freedom that typically are not factored into the interpretation of the results (Gelman & Loken, 2013; Simmons, Nelson, & Simonsohn, 2011).

With complete reporting of all measures, adequate power to find a hypothesized effect, and suitable corrections for multiple testing to maintain a fixed threshold for judging statistical significance (i.e., a fixed alpha level), null-hypothesis significance testing can control for the rate of false-positive results. Correcting for multiple tests provides a measure of error control, keeping the false-positive rate fixed to the alpha (e.g., 5% false positives). Without such corrections, though, the false-positive rate can be substantially higher. Similarly, underreporting of outcome measures makes it impossible for readers to evaluate the likelihood that any reportedly significant result is a false positive because such underreporting means that the alpha level has not been adjusted to account for additional measures. For those reasons, our review highlights cases in which study analyses did not clearly identify all outcome measures or did not correct for multiple testing. Any such evidence must be treated as inherently ambiguous.

Almost all studies in the brain-training literature have relied on null-hypothesis significance testing, but most have not adjusted for multiple tests or tried to maintain a fixed false-positive rate across all measures of the effectiveness of an intervention. Instead, they have treated statistical significance (typically $p < .05$) as a measure of the strength of evidence; the statistical significance of the difference in improvement between the treatment group and the control group is taken as an indicator of how strong the evidence is for the effectiveness of the intervention. Although common, this use of p values is inappropriate; p values can be used to control error rates across a set of related hypothesis tests (assuming appropriate reporting), but individual p values do not constitute a measure of evidence *for* an intervention benefit (see Wasserstein & Lazar, 2016, for a formal statement from the American Statistical Association). They only provide a measure of how unlikely a difference that big or bigger would be if the intervention actually had no effect at all (i.e., that the null hypothesis is true), not how likely those data would be if the intervention were effective. Even if a result is unlikely under the null hypothesis, it also might be unlikely under most plausible alternative hypotheses. Evidence is a relative measure: Which model better accounts for the observed data—one in which the treatment has no effect or one in which the treatment produces a benefit? Null-hypothesis significance testing does not provide such a measure of evidence because it does not assess the relative likelihood of the data under different hypotheses. It assesses the fit of the data to a single hypothesis—in this case, the null hypothesis of no effect.

In fact, a p value just under .05 might be more consistent with the absence of a difference than the presence of one. Imagine an intervention in which you assign 10,000 people to an intervention group and 10,000 people to a control group. If there truly is an effect of that intervention in the population, with a sample that large, pretty much every study would produce a statistically significant result. In fact, the vast majority would produce a *highly* significant result—not one just under .05—even if the true effect of the intervention were small. If the intervention truly had no effect in reality (i.e., if the null hypothesis were true), approximately 5% of studies would produce a significant result with $p < .05$. Under typical conditions, the distribution of p values when the null is true is uniform, so 1% would produce a false-positive result with a p value between .04 and .05, 1% would produce a false-positive result with a p value between .03 and .04, and so on.

Now consider our hypothetical study. We do not know whether the intervention truly has any effect, but we test our 20,000 people and observe a p value between .04 and .05. That would happen about 1% of the time if the null were true. But, perhaps surprisingly, with a sample this large and a true effect—even a small one—we would expect to find a significant result between .04 and .05 substantially less than 1% of the time. In other words, that finding would be more likely if the null hypothesis were true in reality than if there were actually an effect of the intervention. Finding a p value just under .05 provides relatively weak evidence for the presence of an effect relative to the absence of one. And, with large samples, it can actually provide more support for the null hypothesis. More broadly, regardless of the sample size, a p value just under .05 provides *at most* about 2.5 times more evidence for the presence of an effect in reality than for the absence of one (see Bayarri, Benjamin, Berger, & Sellke, 2016). That small a ratio does not provide compelling evidence.

In our review, we try to note cases in which the evidence seems weak (e.g., p values just under .05, small samples, lack of correction for multiple testing). In many cases, the statistically significant results cited by papers on brain-training studies provide at best ambiguous evidence for a benefit (relative to the null hypothesis of no benefit). Ideally, future brain-training interventions will use more appropriate measures of evidence rather than relying exclusively on p values as support for the effectiveness of an intervention, and will shift to the goal of obtaining more precise estimates of the size of any benefits rather than seeking statistical significance. Doing so would make clear the need for larger sample sizes.

Summary

We have identified a set of best practices in study design and reporting for cognitive-intervention trials. Researchers who follow these practices preregister their studies, creating advance documentation of their design and analysis plans as well as all outcome measures. They use appropriately matched active control conditions that attempt to equate for all aspects of the intervention other than the hypothesized critical ingredient, including expectations to the extent possible. Whenever expectations are not equated, the researchers acknowledge the possible role of differential placebo effects. They ensure that the intervention and control groups are comparable prior to the intervention by randomly assigning a large number of participants to each condition. And they measure and equate expectations for improvements on each outcome measure across the intervention and control groups. When publishing the results of the intervention, the researchers report all outcome measures regardless of whether or not they were statistically significant; they adjust for multiple comparisons; and they make clear any overlap between separate papers reporting outcomes from the same intervention.

Unfortunately, no published interventions have yet conformed to all of these best-practice standards. Given that some of these practices, especially preregistration, are relatively new to psychology, we should not expect all brain-training studies to meet all of them. Moreover, in the early stages of intervention research, a lack of adherence to all of these practices is perhaps understandable; exploratory research is a necessary step in identifying hypotheses to be confirmed (or rejected) with more rigorous designs. As long as such preliminary studies explicitly note all of their limitations, note clearly that their findings are exploratory, and do not oversell the implications of their results, less rigorous methods might not be as problematic. Given that many previously promising medical interventions failed when subjected to more rigorous methodological standards (Hay, Rosenthal, Thomas, & Craighead, 2011), findings based on weaker methods should be treated as potentially unreliable and should not be used to guide policy. Furthermore, marketing of brain-training products based on such exploratory evidence is troubling and unwarranted.

In evaluating the quality of the published evidence, we evaluate how well the cited studies meet these methodological, analytical, and reporting standards. We focus our review on those studies that have come closest to meeting these best practices—experimental interventions that compared a cognitive-intervention group to a control group, preferably using random assignment to conditions. When enough information from the study is available, we consider the strength of the evidence for transfer of training from the trained task to other outcome measures.

The Evidence Cited by Brain-Training Companies

Each publication was coded by the first author to determine whether or not it included an intervention, how the intervention was conducted, what tasks were used, and what outcomes were measured. As well as we could on the basis of the published information, we coded the sample sizes, the duration of training, the types of control conditions and training interventions, and whether condition assignment was random so that we could evaluate all papers according to similar standards. When possible, we identified cases in which the results of a single intervention were published in multiple papers. The full coded bibliography is available on the Open Science Framework page for this article at https://goo.gl/N6jY3s. We evaluated the quality of each study by assessing how well it conformed to best practices in intervention research. We also examined the evidence each study provided for benefits on measures other than those used for training.

The open letter from brain-training proponents posted on the Cognitive Training Data website argues that there is a "large and growing body" of evidence for benefits from cognitive training. While acknowledging that individual studies have their limitations, the letter notes that many of the controlled trials show improvements that encompass a broad array of cognitive and everyday activities, show gains that persist for a reasonable amount of time, document positive changes in real-life indices of cognitive health, and employ control strategies designed to account for "placebo" effects. (para. 7)

That statement directly contradicts the earlier open letter from critics of brain training, which claimed that "compelling evidence of general and enduring positive effects on the way people's minds and brains age has remained elusive" and rejected the claim "that brain games offer consumers a scientifically grounded avenue to reduce or reverse cognitive declines when there is no compelling scientific evidence to date that they do."

The question, then, comes down to what constitutes *compelling* evidence. Both open letters acknowledged the need for rigor in conducting brain-training studies. They recognized that studies from independent labs that use active control groups are more compelling, that it is essential to assess both the scope and the persistence of any improvements, and that interventions must control for motivation and expectations. Yet the letters came to incompatible conclusions about the strength of the existing evidence.

Cognitive Training Data provides a list of 132 published journal articles that "directly demonstrate that computerized cognitive training can improve cognition" ("Studies on Cognitive Training Benefits," 2014). Even though this list is said to be incomplete, it presumably constitutes a set of papers that the signatories believe constitute the most compelling support for the efficacy of brain training. We evaluate these citations and then consider additional papers cited by brain-training companies that were not already cited by Cognitive Training Data. Our goal is to determine whether and why that cited evidence is compelling.

In our review, one cognitive intervention merits special treatment, given its scope and impact in the cognitive-training literature: the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) trial (ClinicalTrials.Gov Trial Number NCT00298558). ACTIVE is the largest cognitive-intervention study yet undertaken, rivaling the size of many drug intervention studies: It tested a total of 2,832 older adults (aged 65 or older) across three intervention conditions and a no-contact control condition. This trial has resulted in more than 50 publications, many of which are cited extensively by brain-training companies; Cognitive Training Data includes on its list 15 citations to papers resulting from ACTIVE (see Table 2).

The ACTIVE trial

Data collection in this six-site study of elderly participants began in 1998 (Jobe et al., 2001), with the goal of testing the effects of three distinct 10-hour cognitive interventions relative to a passive control condition. The study examined a wide range of outcomes, including both laboratory measures of cognitive performance and cognitively demanding real-world activities (instrumental activities of daily living [IADLs]) such as food preparation, driving, medication use, and financial management. The ACTIVE study's large sample provided 95% power to detect a differential improvement of 0.20 standard-deviation units, even with a 20% attrition rate. ACTIVE adopted other best practices, including following a published protocol (Jobe et al., 2001) and ensuring that outcome measures were gathered by experimenters who were blind to condition assignment.

Training and outcome measures. More than 2,800 older adults (all aged 65 or older; mean age = 74 years) were randomly assigned to a reasoning training, memory training, speed-of-processing training, or no-contact control

Table 2. The Evidence Cited by Cognitive Training Data

Type of paper	Total number	Percentage
Review or meta-analysis (no new data)	21	16%
Intervention study		
Learning only (no transfer tests)	5	4%
No baseline control group	14	11%
Control group without random assignment	6	5%
Randomized controlled trial		
The ACTIVE trial	15	11%
Passive control group	22	17%
Active control group	49	36%
Total papers cited by Cognitive Training Data	132	100%
Total randomized controlled trials	86	64%

group. Training was conducted by certified trainers who followed scripted training manuals. Each trainer learned to administer only one of the types of training, although backup trainers were familiar with multiple treatment interventions. Participants completed the training in small groups (of 3–5 people) over 10 sessions, each lasting 60 to 75 minutes. Some participants completed the training in a compressed 2-week time frame, but most took 6 weeks. In each training condition, the first five sessions included both cognitive exercises and trainer-modeled strategies designed not just to provide practice with core cognitive abilities but also to boost self-efficacy in cognitive performance and to emphasize how to apply the learned skills and strategies to real-world tasks. The final five sessions included more practice with the cognitive exercises. Throughout training, participants received feedback both individually and as a group. In addition to exploring the immediate effects of training, ACTIVE examined retention of those benefits.

Of those participants who attended more than 80% of the initial training sessions, 60% were invited to return 11 months later for booster training sessions: four additional 75-minute sessions completed over a 2- to 3-week period. These booster sessions resembled the first five training sessions and included both strategy training and practice on cognitive tasks. Additional booster training was offered approximately 3 years after the initial training. In general, the booster training enhanced gains on the trained tasks.

The memory intervention emphasized the application of the mnemonic principles of meaningfulness, organization, visualization, and association to everyday cognitive tasks. The cognitive exercises included both lab-based memory tasks (noun-list recall, paragraph recall) and everyday memory tasks (organizing and recalling items on a shopping list, encoding and remembering details on a prescription label, or remembering a list of errands). Training emphasized episodic memory for word lists, item sequences, text, and the gist and details of stories.

The reasoning intervention used psychometric reasoning tasks, such as serial-pattern and sequence determination, and included both lab-based (e.g., letter-series tests) and everyday activities (e.g., filling a pill-reminder case, making a travel schedule). Given large individual differences in baseline performance, participants in the reasoning training condition were assigned to one of two levels of training based on their initial abilities (Jobe et al., 2001). (It is unclear from the registration how participants were assigned to these two levels, and most reports have not distinguished between these subgroups.)

The speed-of-processing intervention came closest to the sorts of computer-based cognitive tasks commonly used in brain-training products. The training was based on the UFOV (Edwards et al., 2005), which incorporates visual search and divided attention. This adaptive training increased in difficulty as performance improved by shortening presentations, adding distractors (visual and auditory), demanding multitasking, and requiring greater spatial breadth of attention. This adaptive, individualized training contrasts with the group-focused training in the other interventions.

The primary outcomes for ACTIVE (available on the study's ClinicalTrials.Gov page at https://clinicaltrials.gov/ct2/show/NCT00298558; see also Jobe et al., 2001) included measures of memory, reasoning, and speed as well as measures of everyday activities that, collectively, assessed both near and far transfer:

1. Memory performance was measured using the Rey Auditory Verbal Learning Test, the Hopkins Verbal Learning Test, and the Rivermead Behavioural Paragraph Recall test of immediate recall.
2. Reasoning ability was measured using letter-series, letter-sets, and word-series tests.
3. Processing speed was measured using three UFOV tasks. Two other speed-of-processing outcome measures (Digit-Symbol Substitutions and Digit-Symbol Copy tests of the Wechsler Adult Intelligence Scale)

were listed in the baseline paper (Jobe et al., 2001) but were not included in published analyses. It is unclear from the published record whether these additional measures were collected and not reported or were not collected because the original plan changed.

4. Everyday task performance (i.e., IADLs) was measured mostly using self-reports, including the Minimum Dataset - Home Care, reports of performance in the prior week, and reports of the need for assistance in activities of daily living such as dressing and bathing.
5. Everyday problem-solving was measured using the Everyday Problems Test and the Observed Tasks of Daily Living.
6. Everyday processing speed was measured using a complex reaction-time task, a road-sign test, and a timed IADL test.

In addition to these primary outcome measures, ACTIVE included many secondary outcome measures, spanning domains from health quality (health-related quality of life, health-service use) to driving (crashes, driving cessation) to mortality. ACTIVE also collected extensive information about participants, making it possible to examine links between cognitive performance, training interventions, and factors such as disease (cardiovascular disease, diabetes, depression) or the Alzheimer's disease–associated Apolipoprotein E (APOE) genotype.

Results. We base our review of the effectiveness of the ACTIVE interventions on an evaluation of the primary reports of the main outcome measures after 2 years (Ball et al., 2002), 5 years (Willis et al., 2006), and 10 years (Rebok et al., 2014), highlighting those primary outcome measures that were reported consistently across these papers. Overall, the results show a fairly consistent pattern: Improvements for the trained abilities, negligible improvements for untrained abilities, and diminishing effects with increasing time since the training.

Primary outcomes. Table 3 shows the effect sizes reported at each testing stage: immediate posttest, 1-year follow-up, and 2-year follow-up (Ball et al., 2002); 5-year follow-up (Willis et al., 2006); and 10-year follow-up (Rebok et al., 2014). The reported effect sizes for each time of measurement constitute differences between that training group and the no-contact control group in the amount of improvement from the pretest (measured in standard-deviation units).

Immediately after training, each intervention yielded improved performance on laboratory outcome measures of that same ability. Speed training with the UFOV improved performance on UFOV speed-of-processing measures; reasoning training improved performance on letter-series, letter-sets, and word-series measures of reasoning; and memory training improved performance on standard neuropsychological tests of memory. Almost none of the other measures showed differential improvements, and almost none of the training tasks improved performance significantly on the other categories of outcome measures. Speed training improved speed of processing but not memory, memory training improved memory performance but not reasoning, and so on. In effect, each training regimen led to improvements on the trained task, with some near transfer to proximal measures of the same skill but no evidence of transfer beyond the trained task, not even to standardized measures of everyday analogues of the trained abilities. Booster training increased gains on the proximal measures but did not yield increased transfer.

For the speed and reasoning groups, near-transfer effects persisted throughout the 10 years of follow-up testing, although the size of the differential benefits was less than half as large as in the immediate posttest session. The memory effects also persisted at roughly the same size through the 5-year follow-up tests but were largely gone by the 10-year follow-up.

At 5 years after training, the self-reported measure of IADLs showed improvements relative to the no-contact control, and by the 10-year follow-up, IADLs showed modest differential benefits for all three training groups relative to the passive control group. Those benefits were not present at posttest for any of the groups.

The largest training benefits in ACTIVE were for the UFOV-based speed-of-processing intervention. Speed training yielded a differential benefit of nearly 1.5 standard-deviation units on other UFOV measures of processing speed immediately after training, and the benefit was still more than 0.5 standard deviation units 10 years later. Yet speed training appeared to improve performance only for the UFOV-based speed-of-processing measures, those most similar to the training task. It did not produce differential benefits on other speeded tasks, such as choice response time. Memory training produced the smallest differential benefits on proximal measures of memory (about one-quarter of a standard deviation), and reasoning training produced a sizable benefit for reasoning measures (about half of a standard deviation). The larger effect size for speed training than for memory or reasoning training might well have been due to its adaptive, individualized nature or to greater similarity between the trained task and the outcome measures of processing speed.

As expected for an older sample (aged 65 years or older at the start of the study), measures of cognitive abilities mostly showed declines by the 10-year follow-up session (see Table 4). For all groups, performance on

Table 3. Differential Improvements for Training Groups Relative to No-Contact Control Groups in the ACTIVE Trial

Training group and time point	Memory (laboratory measures)	Reasoning (laboratory measures)	Processing speed (laboratory measures)	Self-reported IADL	EPT and OTDL	CRT and TIADL
Memory training						
Posttest	.257*	–0.018	–0.045	NA	NA	–0.091*
Year 1	.212*	0.021	–0.054	.02	-0.045	–0.041
Year 2	.174*	0.045	–0.034	.017	–0.073	–0.007
Year 5	.23*	–0.01	–0.01	.20	–0.15	.04
Year 10	.06	–0.02	–0.07	.48*	.004	.02
Reasoning training						
Posttest	–0.009	0.480*	0.003	NA	NA	0.004
Year 1	–0.011	0.402*	–0.033	–0.125	0.03	0.05
Year 2	–0.03	0.257*	–0.043	–0.056	–0.027	0.03
Year 5	0.05	0.26*	0.15*	0.29*	–0.08	0.09
Year 10	–0.11	0.23*	0.005	0.38*	–0.02	–0.004
Speed training						
Posttest	–0.012	–0.026	–1.463*	NA	NA	–0.016
Year 1	–0.021	–0.003	–1.212*	–0.05	0.008	0.001
Year 2	–0.052	–0.019	–0.867*	–0.07	0.031	–0.009
Year 5	0.05	0.02	0.76*	0.26	–0.05	0.08
Year 10	–0.05	–0.06	0.66*	0.36*	0.008	–0.05

Note: The table shows the effect size reported in the original articles for each type of outcome measure at each testing time point, where effect size was defined as the difference in improvement (relative to baseline) between the experimental and control group, divided by the intra-subject standard deviation. Values are reported with the same number of significant digits as in the original articles. For the immediate test as well as the Year 1 and Year 2 tests, asterisks represent a significantly greater improvement for the intervention group at $p < .05$. For the Year 5 and Year 10 tests, asterisks represent a differential improvement with a 99% confidence interval that excludes zero. For the Year 10 test, the effect sizes were computed using a mixed-effects model. IADL = Instrumental Activities of Daily Living; EPT = Everyday Problems Test; OTDL = Observed Tasks of Daily Living; CRT = complex reaction-time task; TIADL = timed Instrumental Activities of Daily Living; NA = not available.

almost all outcome measures was worse than it had been during the pretest, although declines on tasks related to training were smaller than declines on other tasks. The only exception was for the speed-of-processing outcome measures for the speed-of-processing intervention group, which showed a persisting benefit, on average.

When evaluating training effectiveness, these absolute declines in performance must be judged relative to the declines in the control group, as reflected in the effect-size measures reported in Table 4. Training could be beneficial even if it did not lead to long-term improvements in cognition. Instead, it might effectively limit the extent of the decline in cognitive functioning that would otherwise have occurred.

Secondary outcomes. In addition to the primary outcome measures shown in Table 3, ACTIVE also measured the effects of training on other real-world behaviors. For example, speed-of-processing training was associated with better driving safety, as measured by at-fault crashes. The speed training was based on the UFOV task, which has been shown to correlate with both prospective and retrospective crash risk (Owsley et al., 1998; Owsley, Ball, Sloane, Roenker, & Bruni, 1991). A subset of actively driving participants who could be tracked 6 years after enrollment showed an approximately 50% reduced risk of at-fault crashes (Ball, Edwards, Ross, & McGwin, 2010). Note, though, that the overall crash risk was not substantially lower following training. Brain-training companies often highlight this finding as a key real-world benefit of speed-of-processing training, and we evaluate whether that claim is justified after the literature review, in our "Comparative Effectiveness and Appropriate Inferences" section below. Data from ACTIVE (combined with data from the Staying Keen in Later Life [SKILL] Study) have also been used to investigate whether training predicts driving cessation (Edwards, Delahunt, & Mahncke, 2009), revealing that those in the speed training group were less likely to stop driving. We discuss this analysis when we discuss the SKILL intervention in "The SKILL Study" below. Other secondary outcome measures focused on self-reported health-related quality of life and on depression. Consequently, they fall outside the scope of our review.

Contributions and limitations. The ACTIVE trial has had a large impact on the field. It remains the largest brain-training intervention to date, and it conformed to

Table 4. ACTIVE Trial Training Groups' Mean Changes on Each Outcome Measure Between Initial Baseline Measurement and 10-Year Follow-up

Training group	Memory	Reasoning	Speed	Self-reported IADL	EPT and OTDL	CRT and TIADL
Memory training	−10.6 (28.3)	−3.23 (8.61)	−144 (229)	−3.05 (7.38)	−6.10 (9.75)	−1.53 (2.17)
Reasoning training	−11.2 (26.3)	−0.049 (7.91)	−126 (254)	−2.66 (6.31)	−5.58 (9.56)	−1.39 (1.88)
Speed training	−12.7 (25.5)	−3.94 (8.34)	24.3 (252)	−2.34 (5.62)	−5.98 (9.32)	−1.47 (1.98)
Control condition	−9.4 (29.6)	−3.04 (8.02)	−123 (278)	−3.61 (7.67)	−5.67 (9.85)	−1.42 (1.78)

Note: Each value in this table represents the difference in improvement from baseline to Year 10 for each group and type of measure; these data are drawn from the ClinicalTrials.Gov study record for the ACTIVE trial: https://clinicaltrials.gov/ct2/show/results/NCT00298558. The absolute-improvement numbers are not directly comparable across types of measures because the measure differ in their possible score ranges (memory = 0–132; reasoning = 0–75; processing speed = 0–1,500; IADL = 0–38; everyday speed of processing = −3–100). Most scores declined with age regardless of training group. Standard deviations are shown in parentheses. IADL = Instrumental Activities of Daily Living; EPT = Everyday Problems Test; OTDL = Observed Tasks of Daily Living; CRT = complex reaction-time task; TIADL = timed Instrumental Activities of Daily Living.

many of the best practices for intervention research, including random assignment to groups that were well matched at pretest and careful attention to equating the training regimens in terms of dosing and social engagement. The sample was large, ethnically and geographically diverse (participants were from Alabama, Indiana, Massachusetts, Maryland, Michigan, and Pennsylvania; 28% were non-White), and systematically screened for conditions that might reduce plasticity or interfere with participation in the training (Jobe et al., 2001). Effects of the interventions were examined with both laboratory and real-world outcome measures, with multiple measures tapping each construct, enabling the assessment of effects on latent variables. Participants were followed up for years after the intervention. For each of the major testing sessions, the reporting of the results is a model of clarity, with Consolidated Standards Of Reporting Trials (CONSORT) diagrams (Moher, Schulz, & Altman, 2001) detailing the flow of participants through the study. All primary outcomes at each testing time were reported for the whole sample in a single paper, and those papers reported the reliability of their measures.

The results of this large study were clear in showing effectively no evidence of transfer of training from one skill to another; improvements were mostly limited to the trained activities. The speed-of-processing training group showed the largest gains both immediately and over time, but the benefits were limited to similar outcome measures of processing speed and did not transfer to other speeded responses. The benefits in this group (relative to the control group) also diminished over time.

Of the training interventions, speed of processing is arguably the most likely candidate to produce far transfer. Speed of processing contributes to many higher-order cognitive processes (Salthouse, 1996, 2010), and it is a rate-limiting factor for both memory (e.g., encoding speed for perceptual information, rehearsal rate, motor response speed) and problem-solving (generating and traversing states in a problem space). Although speed of processing likely consists of multiple distinct processes (Hale, Myerson, Faust, & Fristoe, 1995; Jastrzembski & Charness, 2007), faster visual-search processes might be expected to help in other tasks that require fast decisions and reactions. Yet ACTIVE showed little evidence that speed training improved performance on other cognitive tasks.

Despite its contributions, ACTIVE had limitations. For example, the no-contact control condition was not an ideal baseline against which to compare the effects of training, and almost all of the observed benefits were relative to that control condition. Had the trial more systematically compared the different training conditions directly, each could have served as a control condition for the other groups (Willis, 2001), albeit one not necessarily matched for differential expectations. The scope of the study itself introduces some complications for interpreting its results. For example, ACTIVE is one of the few studies to have tracked participants for 10 years after an intervention. But because participants completed each of the outcome measures at each testing session, the study likely underestimated the age-related declines that might have occurred without this repeated practice.

After 10 years of testing reported across more than 50 papers, evaluating the strength of any individual finding becomes difficult. Aside from the key reports of the results from all participants on the primary outcome measure after each testing period, other papers report a variety of outcomes for subsets of the participants (e.g., Ball et al., 2010). How can these separate reports be adjusted to correct for multiple comparisons? The lack of correction increases the likelihood that some of these additional findings are spurious. It is important to keep in mind that the 50-plus papers resulting from the ACTIVE trial all reported results from just one study, and they should not be regarded as independent pieces of evidence for training benefits. That is true even for the key results papers from each time period—they report outcomes from the same original training groups, not from new groups of participants.

Unlike many cognitive interventions, ACTIVE measured meaningful real-world outcomes. For older adults, the ability to maintain independent living is perhaps the most societally important benefit of cognitive training. Unfortunately, the outcome measures that most directly tap such benefits (IADLs) are also the least objective of the outcome measures, and such benefits were not present for more objective measures of daily performance. Because they are based on self-reported behavior rather than direct observation, they are more vulnerable to expectancy effects (e.g., Cortese et al., 2015; Rapport, Orban, Kofler, & Friedman, 2013). However, for people showing difficulties with IADL activities, self-reports do tend to predict highly significant real-world outcomes, including the likelihood of institutionalization (Luppa et al., 2010). It is possible that these benefits took time to emerge and that they represent long-term benefits of training on real-world performance.

The IADL results are somewhat puzzling, though, because all three intervention groups showed benefits relative to the no-contact control group at the 10-year testing point. Given that the three intervention conditions trained different aspects of cognition, it is unclear why they would produce comparable benefits for IADLs. However, all three interventions also included other components, including strategy training, opportunities for social interaction, and self-efficacy training, and it might be these aspects of the ACTIVE intervention, rather than the cognitive content of the training, that enhanced self-reported IADLs. Consistent with that explanation, 10 years after the intervention, memory training produced no benefits on the trained tasks but still enhanced IADLs. It is unusual to find a case in which a cognitive intervention improved underlying mechanisms in a way that yielded far transfer to everyday activities but not near transfer to highly similar tasks. Whether the increased motivation to perform well in daily life could yield greater transfer benefits or whether self-report measures simply reflect the results of expectations and experiment demands can be hard to determine without additional objective measures of performance.

Summary. In many ways, ACTIVE is a sound model for a randomized clinical trial assessing the impact of a cognitive intervention. Among its strengths are the large sample size, the long follow-up period (10 years), and the publication of a "design" paper before the first report of outcome measures (Jobe et al., 2001). Its weaknesses include a non-active control group (no-contact control), the failure to measure and compare participant expectations across conditions, and publications with subgroup analyses that may increase the risk of type I error. Nonetheless, future intervention studies can benefit from a careful study of ACTIVE's design and methodology. The results were clear in showing effectively no evidence of transfer of training from one skill to another; improvements were limited to the trained skills, and there was no compelling evidence for far transfer to other types of tasks.

The evidence cited by Cognitive Training Data

Cognitive Training Data cites 132 journal articles in support of the benefits of cognitive interventions. We first classify the studies reported in these publications (see Table 2) and then evaluate the quality of the intervention studies (see https://goo.gl/N6jY3s for our coding of each article in this set).

Of the 132 Cognitive Training Data citations, 21 are review articles or meta-analyses that presented no new data and five explored improvements on the training task without reporting whether training transferred to other outcome measures. The remaining 106 reported interventions that included outcome measures. Of those, 15 reported findings from the ACTIVE study discussed earlier (recall that ACTIVE provided minimal evidence of transfer beyond the trained abilities). Another 14 papers lacked a baseline comparison group. With that design, any transfer of training might result from many factors other than the hypothesized critical ingredient in the training itself (e.g., perhaps most significantly, test-retest effects, but also social contact and placebo effects). Consequently, such studies cannot be taken as evidence for the effectiveness of brain training.

Six papers included a control group but did not randomly assign participants to conditions. Three of these explicitly noted that participants were matched across the experimental and control groups rather than randomly assigned. One tested the effectiveness of a cognitive intervention for depression symptoms and compared improvements to those of a "historical" control group whose participants had previously received antidepressant drug therapy (Morimoto et al., 2014). Another used blocked recruiting in which participants were assigned first to one condition and then to the other (e.g., Haimov & Shatil, 2013). Finally, one partially blocked and partially matched assignment to conditions based on lesion locations (Van Vleet & DeGutis, 2013). Although matched or blocked control groups can provide a useful baseline, they do not equate for possible cohort effects. Three other papers used control groups drawn from other studies or made clear that participants were not randomly assigned to conditions.

The remaining 71 papers reported the results of an intervention in which participants were randomly assigned to conditions. (That is, the papers did not explicitly state that assignment was *not* random; by default, we assumed that assignment to conditions was random unless the methods made clear that it was not. If we were incorrect in this assumption, our conclusions might overestimate the effectiveness of the interventions because of possible biases in assignment to conditions.) In a handful of these cases, participants were randomly assigned to multiple treatment groups without a clearly defined baseline group. When the effectiveness of a training intervention must be compared to that of another intervention rather than a baseline condition, it is possible that any benefits could result from factors other than the critical ingredients of the intervention. For example, if both treatment groups improved on the same outcome measures, the improvements in both cases could have resulted from a placebo effect, from social contact, from test-retest effects, or from a number of other factors. Without a baseline condition, such studies provide evidence only for differences in the effects of the interventions and do not provide evidence independent of the effects of retesting or expectations. Moreover, they might mask actual training benefits: If both training groups improve to the same extent, it is possible that both interventions are effective or that neither one is.

These 71 published reports of intervention studies did not, however, constitute 71 independent intervention studies, because results from several interventions were reported in multiple papers. Several large-scale interventions (in addition to ACTIVE) accounted for multiple cited papers: the SKILL trial (five papers, one of which also analyzed data from ACTIVE), the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) trial (three papers), and the Iowa Healthy and Active Minds Study (IHAMS; three papers). In addition to these "named" trials, several additional interventions accounted for a large number of published papers. These included two interventions for cognitive rehabilitation in schizophrenia (ClinicalTrial.Gov Trials NCT00312962 and NCT00430560), which accounted for 11 and seven of the cited papers, respectively.

Many of the remaining interventions produced multiple published articles cited by Cognitive Training Data as well. Some did not clearly acknowledge that they were reporting data from the same intervention. Some did not cite the clinical trial number; others reported outcome measures without noting that the participants were drawn from the same intervention that had been reported in a separate paper; and still others reported subsets of the data from a larger intervention. When the reported evidence across multiple papers was drawn from a single intervention, we treated those papers as non-independent and reviewed them as a single study. Often we had to infer overlap from partial or incomplete reporting. We did so based on overlap in the team of authors, the use of identical interventions, testing of the same subject populations, testing during the same time period, or similarities in the data or training procedures.

Of the 71 published papers reporting the results of randomized trials (aside from ACTIVE), 22 compared the treatment group(s) to a passive control group—one in which participants did not engage in any tasks other than their regular activities (e.g., no-contact, waitlist, or treatment as usual). A total of 49 papers reported the outcome of an intervention in which the treatment group was compared to an active control group—one that engaged in some experimenter-assigned activity over the time period of the intervention. In almost all cases, these active control groups spent the same amount of time on their control tasks as the intervention groups did on theirs, thereby equating non-experimental factors such as social contact that could contribute to improved performance.

Cognitive Training Data evidence from interventions with passive control conditions. Cognitive Training Data cites 22 passively controlled interventions (other than ACTIVE). Three of these examined the effects of cognitive training on balance and gait, measures known to be related to the risk of falling (Smith-Ray et al., 2013; Smith-Ray, Makowski-Woidan, & Hughes, 2014; Verghese, Mahoney, Ambrose, Wang, & Holtzer, 2010), but they did not report the effects of cognitive training on cognitive performance, so they fall outside the scope of our review. The 19 remaining papers included seven from Clinical Trial NCT00430560, and several other interventions resulted in multiple papers as well. As for ACTIVE, we treat each collection of papers as a unit.

Clinical Trial NCT00430560. This clinical trial resulted in many papers, including the seven cited by Cognitive Training Data. It explored the effects of cognitive training on cognition, work behavior, and clinical symptoms for people with schizophrenia. The registration itself was not posted until January 2007, after the completion of the intervention, which explains why a number of the papers reporting on this study did not cite the clinical trial number (of the cited papers, only Bell, Fiszdon, Greig, Wexler, & Bryson, 2007, does). According to the registration, primary data collection took place between 2000 and 2005 (although some papers reported recruiting between 1998 and 2003). The registration estimates enrollment of 150 participants in total, with random assignment of matched pairs of participants to a work-therapy control group or a work-therapy-plus-cognitive-training intervention group, but it provides no details about the training, testing, or specific outcome measures.

The papers resulting from this study noted that the intervention group received standard therapy and work training along with both cognitive and social training. The cognitive-training materials consisted of packages from Scientific Learning Corporation (developed with the company by one of the authors, Bruce E. Wexler) and Psychological Software Services (CogRehab). Some of the cited papers focused on improvements in the training task (e.g., Fiszdon, Bryson, Wexler, & Bell, 2004), and others were review articles. Here, we focus only on those that reported analyses of transfer to cognitive or workplace outcome measures from the intervention.

The first paper in this set (Bell, Bryson, Greig, Corcoran, & Wexler, 2001) reported the results of a 5-month (approximately 130-hour) cognitive intervention with 31 patients with schizophrenia in the intervention group and 34 in the control group. The outcome measures consisted of a battery of 15 standardized neuropsychological tasks as well as questionnaire measures of work behavior, cognition in the workplace, and clinical symptoms. The training consisted of social therapy, cognitive therapy, and biweekly cognitive feedback based on job performance. This paper reported training with five cognitive measures adapted from CogRehab and designed to tap attention, memory, and executive functioning. Thus, the trained tasks were designed to train the same cognitive skills measured by the outcome tasks. In some cases, the tasks were highly similar (e.g., digit recall at training and digit span at test).

The primary analysis focused on composite measures created based on factor analysis. Of the five factors, two were significant, but only one (consisting of only the Wisconsin Card Sorting Test) would have been significant after correction for five comparisons. Subsequent analyses separately examined each outcome measure. Of the 21 identified measures, five showed differences between the training and control groups at posttest when controlling for pretest scores, but three of those five would not have been significant at $p < .05$ with correction. The intervention provides some evidence for improvement on the Wisconsin Card Sorting Test, but given the nature of the intervention, it is unclear whether these benefits constitute narrow or broad transfer. If they do constitute transfer, whether those improvements were due to the cognitive training, the social training, or the cognitive feedback is unclear. That issue applies to all of the papers in this collection.

Bell, Bryson, and Wexler (2003) reported on the same study after data from 102 participants had been collected. This paper focused on the backward-digit-span outcome measure and did not report any of the other outcome measures examined in 2001. The absence of reporting of the other measures with the larger sample raises the possibility that they were no longer statistically significant. If so, the paper reporting on the smaller sample should ideally be corrected to note that it is superseded by the larger, later report. The analyses split the sample into participants with more or less severe cognitive impairments and conducted the primary analyses incorporating level of impairment. It reported results for the 6-month and 12-month follow-up stages but did not reevaluate the effects immediately after training. These changes in analytical approach make a direct comparison between the results reported in this paper and the 2001 paper difficult. The outcome measure discussed in this paper (digit span) was highly similar to one used in training, so any benefits reflect only near transfer to a similar laboratory task.

Bell, Bryson, Greig, Fiszdon, and Wexler (2005) focused on work outcomes at 6- and 12-month follow-up points, reporting a larger sample than either of the earlier papers. The work outcomes consisted largely of hours worked and pay received for continued work after training. Unfortunately, the paper did not report any of the cognitive outcome measures with that larger sample. The study found evidence that by the 12-month follow-up, participants who had received cognitive training in addition to work therapy had worked more hours. However, the means were in the opposite direction at 6 months after the intervention—those who received no cognitive training worked more hours. The Time × Condition interaction was not significant for the outcome measure of dollars earned. Thus, the study provides ambiguous evidence about the extent to which cognitive training enhanced work success. Bell, Zito, Greig, and Wexler (2008a) divided participants into those with higher or lower levels of community functioning and observed bigger workplace gains for those who were initially lower functioning. Bell, Zito, Greig, and Wexler (2008b) conducted additional analyses of the 12-month work outcomes, largely confirming the results of Bell et al. (2005).

Bell et al. (2007) reported the neuropsychological outcome measures for a sample of 53 participants in the intervention group and 63 in the control group, but only after a 6-month delay following the pretest. The authors did not report the same outcome-measure tests immediately after training with their larger sample, instead reporting linear trend contrasts for each of the outcome measures. Given that we should expect the benefits of training to weaken with an increasing delay, a linear trend would not be expected, so it is unclear why that analysis was chosen. Some of the same measures as in the 2001 paper showed differential improvements between groups, but the size of those differences with this larger sample is unclear. Again, the few significant results were not corrected for multiple comparisons.

A final paper in this set (Greig, Zito, Wexler, Fiszdon, & Bell, 2007) reported outcomes from a smaller sample of participants (33 in the intervention group, 29 in the control group) in a 1-year follow-up analysis. The paper reported significant differences for three cognitive tasks, but two would not have been significant after correction for multiple comparisons (and one was misreported). Overall, the pattern for this study at the 1-year follow-up point was comparable to that for the original 2001 paper.

Taken as a collection, these papers provide minimal evidence for transfer of training, with some hints that effects found with partial samples were not present with larger samples. Unfortunately, none of the later papers with larger samples reported the same outcome analyses as the earlier ones, making a direct comparison impossible. The effects on cognitive outcome measures that were observed are consistent with narrow transfer, with benefits persisting in follow-up testing. The evidence for workplace benefits resulting from a combination of work therapy and cognitive training were mixed, with a benefit emerging only for the number of hours worked at the 12-month follow-up (but not the 6-month follow-up) and no significant benefit for dollars earned. Given the lack of consistent reporting across papers, it is difficult to determine whether this intervention yielded any evidence that training enhances performance on skills other than the trained ones.

Other interventions reported across multiple papers. In addition to Clinical Trial NCT00430560, several other interventions with passive control groups resulted in multiple papers cited by Cognitive Training Data. Two potentially related papers examined the effects of Posit Science's InSight software, a speed-of-processing regimen based on the UFOV. Neither paper mentioned the testing reported by the other paper, although they used the same training regimen for the same amounts of time and with the same subject population (although sample sizes differed). Edwards, Valdés, et al. (2013) reported greater improvements in UFOV outcome measures for participants receiving speed training. O'Brien et al. (2013) examined event-related potentials (ERPs) during a visual-search task, finding no difference between the training and control groups in behavioral performance. Neither study observed differential improvements on outcome measures other than the UFOV. Collectively, these studies provide no evidence for transfer of training beyond the trained skill. It is unclear whether any other measures were collected or tested.

Another pair of papers examined the impact of 10 hours of visual-discrimination training (Sweep Seeker from Posit Science's InSight program) on a similar motion-discrimination task and on working-memory performance in older adults (Berry et al., 2010; Mishra, Rolle, & Gazzaley, 2014). Both outcome measures showed differential improvements. The motion-discrimination task tapped the same skill as the training task, and the memory outcome measure used the same stimuli that were used in training. It is unclear whether memory tasks that differed in content would show improvements. The transfer to a memory task would constitute an example of relatively far transfer via improved sensory processing. Further research should attempt to replicate this finding with a larger sample, an active control condition, and varied stimuli to determine whether the findings are robust and if they are limited to improved processing of those stimuli.

Individual papers reporting all results from an intervention. In addition to these collections of papers, Cognitive Training Data cited other publications that appear to be the sole reports of the results of interventions that used passive control groups. Although it is possible that other papers not cited by Cognitive Training Data also reported outcomes from some of these interventions, we found no direct evidence for such overlap among the papers cited by Cognitive Training Data or other brain-training companies.

Like ACTIVE, several of the studies reported in these papers examined the benefits of speed training using variants of the UFOV. Edwards et al. (2002) examined the effects of 10 hours of training on 16 cognitive measures in 44 elderly adults. Of those 16 measures (without correcting for multiple comparisons), only the UFOV and a timed IADL task showed a differential benefit of speed training (controlled oral word association improved more in the control group). In a pilot study, Vance, Fazeli, Ross, Wadley, and Ball (2012) found the same pattern for 22 older adults with HIV, with bigger improvements for the speed-training group on the UFOV and timed IADL tasks and bigger improvements for the control group on a finger-tapping task. Some of these improvements would not have been significant had the analyses corrected for multiple comparisons. Six of the studies cited by Cognitive Training Data, this one included, were pilot or feasibility studies. The primary purpose of a pilot study is to refine measures and procedures, not to conduct confirmatory hypothesis tests about the effectiveness of an intervention (Leon, Davis, & Kraemer, 2011). Such studies should be interpreted with caution.

Von Ah et al. (2012) compared the effects of speed training to those of memory training for breast cancer survivors. The memory training group improved more on a composite memory measure, but only after a 2-month delay, not immediately after testing. The speed training group showed improved UFOV performance and immediate memory at both time points and improved delayed memory at the 2-month follow-up (but not immediately). However, these analyses were contingent on controlling for age and education differences between groups, and none of the near-transfer differences would have been significant had the tests been corrected for multiple comparisons. Thus, the study provides little evidence for near transfer but some inconsistent evidence for a benefit of speed training on memory performance. Both training protocols also led to some improvements on self-report measures of cognitive functioning.

Finally, Scalf et al. (2007) trained 25 older adults for approximately 4 hours using the Functional Field of View (FFOV), a free variant of the UFOV that measures attention breadth and processing speed. Although performance improved on the trained task, those improvements did not transfer to another version of the same FFOV task, which showed that attention breadth was not improved more broadly. Although the study collected a larger battery of cognitive measures, they were used only to explore differences between the groups at baseline, and no measures of transfer of training were provided. In short, the study provided no evidence of transfer of training.

In addition to the papers examining speed training, two cited studies tested the effects of training on a large battery of tasks, one using software from CogniFit (Preiss, Shatil, Čermáková, Cimermanová, & Ram, 2013) and another using Dakim Brain Fitness, a brain-training program with more than 400 exercises across six target domains (K. J. Miller et al., 2013). Preiss et al. (2013) trained 24 patients with unipolar depression for 8 hours and tested improvements on seven cognitive constructs, finding differential improvements for global executive control, shifting, and divided attention. However, none of these effects would have been statistically significant at $p < .05$ after correction for multiple comparisons. K. J. Miller et al. (2013) tested 36 older adults after approximately 20 and 40 hours of training. After 20 hours, the control-group participants were also allowed to play the training games, so only the 20-hour testing session had a baseline comparison group. Of the many outcome measures, only a delayed-memory composite measure differed in a Time × Condition interaction, but the effect is difficult to interpret because it included all three time points in the analysis. Thus, neither study provides clear evidence for transfer. Moreover, without more complete reporting of the similarities between the myriad trained tasks and the outcome measures, it is not possible to determine whether any observed benefits constitute transfer or just improvements on the trained skill.

Two papers explored other forms of training on patient populations. DeGutis and Van Vleet (2010) examined the effects of 5.5 hours of training with a variant of a continuous-performance task in which participants responded as quickly as possible to most images but withheld their response to a target image. (Note that assignment to conditions in this study was not entirely random: The first 4 participants to enroll received training, the next 13 participants were randomly assigned to conditions, and the final 7 participants were assigned to the control group.) The trained group showed an advantage in a conjunction-search task (although the analysis did not include a test of the Time × Condition interaction alone), improved detection performance in the attentional blink, and a reduction in the spatial bias that typically accompanies neglect. Whether these improvements reflect a change in core abilities or a change in bias/perseverance is unclear. Both may be useful for neglect patients, although the improvements seem relatively short-lived and it is unclear if they generalize beyond these laboratory tasks. The paper included a second study of 3 of the same patients who received visual-search training, but that study lacked a control group.

Finally, Polat et al. (2012) examined the effects of approximately 20 hours of visual-detection training on 30 people with presbyopia. The authors compared improvements on various measures of perception (acuity, reading speed, contrast detection, and contrast discrimination) in relation to performance by a baseline group of 3 people with presbyopia who did not undergo training and a second baseline group of 7 young controls with no vision impairments. Although the training group showed promising improvements in most of the psychophysical measures of perception and performance and the control groups did not, the paper did not report the Time × Condition interaction (with a control sample of 3 participants, such an analysis would be underpowered to detect anything but the largest effects). The approach of training basic sensory performance—something that could potentially affect a wide range of behaviors—is promising, but the study itself did not provide a direct test of the effectiveness of training relative to a suitable control (see Jacoby & Ahissar, 2015, for a discussion of the lack of appropriate control groups in visual-learning intervention studies).

Summary: Cognitive Training Data evidence from studies using passive control groups. Collectively, the results from interventions using passive control groups are consistent with the results from ACTIVE. Training typically improved performance on the trained tasks or close variants but did not transfer to cognitive tasks that tapped other abilities. Most of these studies tested small samples, and many tested patient populations with cognitive deficits (e.g., people with schizophrenia or depression, people receiving cancer treatment, people suffering from spatial neglect or presbyopia). Studying the remediating effects of cognitive training is worthwhile, but it is not clear whether any benefits observed for those populations would generalize to the broader public or if they would help with more typical cognitive aging.

Many of the cited papers are part of collections reporting different outcomes from a single intervention. Consequently, the quantity of evidence for training benefits is not as large as implied by a count of the number of publications. Moreover, none of these papers adequately corrected for multiple comparisons, and many of the results reported as significant would not have been after correction. Some papers did clearly identify all of the outcome measures collected, but others did not report statistical tests that would allow a clear determination of whether the intervention group outperformed the control group.

Even if these studies had provided evidence for significantly greater improvements in the intervention group than in the control group, the use of a passive control group limits the strength of the conclusions we could draw. Passive control groups do not adequately control for differential expectations, motivation, experimenter demand effects, or many other possible factors that could cause an intervention to yield bigger improvements. In sum, these studies provide little compelling evidence for the efficacy of brain training as a way to induce broad cognitive improvements.

Cognitive Training Data evidence from interventions with active control conditions. Cognitive Training Data cited 49 papers in which an intervention group was compared to an active control group. Of those, approximately half came from a total of six intervention studies: three from IHAMS, five from SKILL, three from IMPACT, 11 from Clinical Trial NCT00312962, and three each from two unnamed interventions. Some of the remaining 25 papers might overlap in ways that could not be determined from their reporting.

We first consider collections of papers resulting from a common intervention study, and we then examine papers that appear to be the sole report of an intervention. We exclude from consideration papers that only analyzed improvements of the training tasks or that examined the treatment group without comparing its performance to that of the control group.

The IHAMS study (Clinical Trial NCT01165463). The IHAMS trial is the closest any intervention has come to providing a direct replication of the ACTIVE trial, in terms of both its sample size and the intervention content (specifically, the speed-of-processing arm of ACTIVE; Wolinsky, Vander Weg, Howren, Jones, & Dotson, 2013; Wolinsky et al., 2011). This study was substantially larger than most other cognitive-training studies, with 681 participants across four conditions. IHAMS was designed to address a number of limitations of the ACTIVE trial, including the use of a no-contact control group. It also expanded the age range of the included participants (50 years and older, compared to 65 years and older in ACTIVE).

Participants in the intervention groups trained on a variant of the UFOV that is part of the "Road Tour" exercise in Posit Science's DriveSharp program (Posit Science Corporation, 2010) and completed their training either in the lab or at home (for 10 or 14 hours, depending on whether they were assigned to a booster training group). In the task, participants indicated the location of a peripheral object (in this case, a Route 66 sign) while simultaneously judging whether a vehicle presented at the center of their gaze was a truck or a car. The difficulty of the task increased throughout training in that (a) the cars and trucks became more similar to each other, (b) peripheral distractors were added, (c) backgrounds became more complex, and (d) road signs moved farther into the periphery. Moreover, as participants improved, the presentation durations decreased.

The primary outcome measure was performance on a similar UFOV task in which participants localized a peripheral car target among triangles while identifying a centrally presented object as a truck or a car. In effect, this outcome measure is more an assessment of training gains than transfer of training because it tests the same skills using almost identical materials. Secondary outcome measures included a number of other measures of processing speed, attention, and executive control (see Wolinsky et al., 2011, for the study protocol). These included standard neuropsychological measures such as Trail Making Test A and B and the Symbol Digit Modalities Test.

One paper reported the results for the primary outcome measure immediately after training (Wolinsky et al., 2011), and another reported transfer effects 1 year after training (Wolinsky et al., 2013). Immediately after training, participants trained on the UFOV showed greater improvements on the UFOV outcome measure, a benefit that persisted 1 year later. Given the similarity between the training task and this primary outcome measure, this result can be considered a validation that training improved performance on what was effectively the same task.

Improvements on the secondary outcome measures were not reported in the paper describing the immediate effects of training, but they were reported for the testing session 1 year later. At that point, small but significant transfer effects were observed for four out of six transfer tasks ($p < .05$). Although the authors corrected for the multiple tests necessary to compare each treatment group to the control group, it is not clear that they corrected for the number of ways that these outcome measures could have been analyzed (e.g., nine performance metrics can be derived from these six transfer tests). Moreover, the control group had significantly worse baseline performance than the intervention group for the majority of the secondary outcome measures (Wolinsky et al., 2011). That baseline difference raises concerns that random assignment did not control for important group differences, and it potentially undermines the validity of comparisons across groups.

Although IHAMS improved upon ACTIVE by using an active control group, the chosen control task (crossword puzzles) was not well matched to the demands of the intervention. Participants might have expected greater improvements on measures of speed and attention from an intervention that also emphasized speed. The intervention also might have caused strategy changes (e.g., shifting the relative emphasis participants gave to speed and accuracy) that could change performance even if the underlying abilities were unchanged by the intervention. In combination with the different levels of baseline performance between the intervention and control groups, the lack of a well-matched control task potentially undermines conclusions about transfer of training.

The SKILL study. The SKILL study examined the effects of speed-of-processing training, focusing only on older adults already exhibiting processing-speed deficits as assessed using subtests of the UFOV. Edwards et al. (2005) reported the results of a study using a subset of participants from the SKILL sample who were randomly assigned to receive processing-speed training or Internet training (n = 63 for each group). The speed-of-processing training program was not fully described, but it included elements of the UFOV: computerized training in which participants identified central targets and/or localized peripheral targets. The difficulty of the training task was increased throughout training by reducing display times and adjusting the complexity of the central and peripheral tasks. This training group also participated in discussions of mobility, driving, and how speed of processing was important in everyday activities. The control group was taught how to access websites and use e-mail but received no cognitive training. The primary outcome measure was a variant of the UFOV similar to that used in the IHAMS study: Participants identified a central object as a car or truck and indicated the location of a car in the periphery. Secondary outcome measures included tasks designed to measure executive function, speed, memory, and everyday performance (e.g., the timed IADL, which measures how long it takes to perform simple tasks such as looking up a phone number in a phone book or finding information on a canned food label).

After 10 hours of training, two out of 11 outcome measures demonstrated larger gains for the intervention group than for the control group (UFOV, timed IADL tasks). To guard against the increased likelihood of spuriously significant differences that might result from conducting separate analyses for each outcome measure, the authors first conducted a multivariate analysis of variance (MANOVA) and observed a significant Time × Condition interaction. Unfortunately, doing so does not protect against the problems with multiple comparisons. Given the similarity of the training task to the primary outcome measure, a MANOVA that included the UFOV along with the transfer tasks would effectively guarantee a significant interaction effect if training was at all effective in improving performance on the trained task. Had the UFOV not been included in the MANOVA, the interaction effect might well have been nonsignificant. And, had the individual analyses of each outcome measure been corrected for multiple comparisons, the one measure indicating transfer (IADL; p = .028) would not have survived correction. Consequently, the one fairly robust differential improvement in this study was for a measure that tapped the same skills as the training tasks and used highly similar stimuli. This finding provides evidence for an improvement on the trained task or a highly similar one (near transfer) but not transfer to a different cognitive skill or task (far transfer). As for the IHAMS study, the speed-training intervention was adaptive, but the control-group training was not. The two conditions likely induced different expectations for improvement on the outcome measures, and they might also have induced different strategies.

The SKILL sample was also used to examine whether training affects performance outside of the laboratory. Three years after the intervention, all participants provided self-reports of their driving mobility (Edwards, Myers, et al., 2009). Recall that participants in the training and control groups had already experienced performance problems on the UFOV. In this secondary analysis of real-world outcomes, their performance was compared to that of an unimpaired reference group (n = 366) under the assumption that the trained group should show performance comparable to that of this unimpaired group, but the control group should show steeper declines. Changes over time for the speed training group were not significantly different from those for the unimpaired reference group, but they were significantly different for the control group. However, these effects were small, the analyses included many covariates that did not differ significantly between groups, and the results were not reported without those covariates. Moreover, the paper did not report a direct test of the difference between the intervention and control groups, and only that comparison would allow a clear inference of a training benefit. Consequently, the reported results do not provide compelling evidence for a benefit of speed-of-processing training on self-reported driving mobility.

A related analysis combined data from both the SKILL and ACTIVE studies to explore whether speed training would help older adults to keep driving (Edwards, Delahunt, & Mahncke, 2009). The logic of this test followed from the assumption that speed training on the UFOV would enhance driving performance, which in turn would allow older drivers to remain fit to drive. Based on self-reported driving cessation, those participants who attended at least eight of the 10 training sessions were less likely to stop driving over the subsequent 3 years if they were in the speed-training group compared to the control group (Edwards, Delahunt, &

Mahncke, 2009); 9% of the speed-trained group stopped driving, compared to 14% of the control group. Note, though, that this analysis depended on excluding participants who attended fewer than eight sessions, and the results were not significant in an intent-to-treat analysis. The choice to focus on those participants who completed enough training is reasonable—we would not expect improvements from those who did not experience the intervention—but it raises concerns about the robustness of the effect.

Moreover, without objective measures of driving performance, we cannot determine whether staying on the road longer is a positive consequence of the training regimen. The decision to stop driving is a complex one, influenced by many factors beyond driving skill. The inclusion of discussions of how to apply speed training to activities such as driving and mobility might have increased expectations that the intervention would help driving. If so, the intervention might have increased confidence about driving even if it did not improve driving performance. Perhaps confidence led more of the trained participants to keep driving because they thought that they were better drivers, even if they actually were not. If this speculation proves true, and training did not objectively improve driving performance, then having a higher percentage of participants continue driving would be an adverse outcome of the intervention.

The IMPACT study. The IMPACT study compared the effects of 40 hours of Posit Science Brain Fitness auditory training to 40 hours of watching educational DVDs and answering quiz questions. The primary outcome measures included a set of standardized cognitive tasks known as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), focusing especially on those that used auditory presentation/testing of attention and memory. The study including a number of additional measures of learning and memory as well as survey measures of cognitive abilities, cognitive failures, and depression. It tested nearly 500 older adults, making it one of the larger brain-training interventions using an active control condition. The design of this study was more rigorous than most, with careful selection of participants to exclude those with signs of cognitive impairment or dementia, a large sample, appropriate blinding to conditions, and an active control group. Cognitive Training Data cited three papers reporting outcomes from this study: Smith et al. (2009) reported the primary outcomes, Zelinski et al. (2011) reported the results of a 3-month follow-up, and Zelinski, Peters, Hindin, and Petway (2014) re-analyzed the original data using structural equation modeling to look for relations among the tasks. (Note that the IMPACT study was financed by Posit Science and included company employees and consultants as authors, so it is not an independent test of the product's effectiveness.)

Smith et al. (2009) reported significantly greater improvement in the auditory memory/attention composite from the RBANS for the training group than for the control group. The training group also showed greater improvements on an overall memory measure, backward digit span, letter-number sequencing, Rey Auditory Verbal Learning Test total score, and delayed recall. No differential improvements were observed for the Rivermead Behavioural Memory Test (immediate or delayed). These analyses were not corrected for multiple comparisons, and several would not have been statistically significant after correction.

With large sample sizes, a true effect in the population will, on average, result in smaller p values, not p values just barely under .05. In fact, p values just under .05 could even be more likely if the null hypothesis of no difference were true than if the alternative of a positive effect were true (e.g., Lakens & Evers, 2014). Zelinski et al. (2014) computed Bayes factors for these outcome measures, an index of the relative evidence for the null hypothesis and the hypothesis of a difference between groups in improvement scores, and found that one of the significant effects provided stronger evidence for the null hypothesis than for an advantage for the intervention group. And the average Bayes factor across nine outcome measures reported in Smith et al. (2009) provided minimal evidence for transfer.

In short, this large intervention found evidence for greater improvements on auditory measures of learning and memory following 40 hours of auditory perception and memory training than following time spent watching educational DVDs. Those effects that would be robust to correction for multiple comparisons might be considered examples of narrow transfer, given that they relied mostly on other auditory perception/memory tasks that tapped constructs fairly similar to those of the training task. Participants in the training group also self-reported better cognition, although it is unclear whether such self-reports translate into better objective cognitive performance outside of the laboratory.

Zelinski et al. (2011) reported the results of a 3-month post-training follow-up that showed a similar pattern of results, with somewhat reduced effects. This paper used a more sophisticated analysis strategy, including mixed-effects modeling and ANCOVA controlling for baseline performance when comparing performance on the follow-up outcome measures across groups. Even after 40 hours of training, the primary outcome measure showed no benefits when assessed again just 3 months later. The effects on overall memory, processing speed, Rey Auditory Verbal Learning Test total score, and letter-number sequencing remained significant at the 3-month follow-up. In most cases, the effects at follow-up were slightly smaller than immediately after training. These findings suggest that training benefits were fairly persistent.

The final paper in this set (Zelinski et al., 2014) reanalyzed data reported in earlier papers in an effort to control for the relations among tasks when looking at the correlation between training gains and outcome-measure improvements. Such correlations between training improvements and outcome-measure improvements are often taken as evidence for transfer of training, but they likely are not a valid way to determine whether or not training transferred to the outcome measures (Jacoby & Ahissar, 2015; Tidwell, Dougherty, Chrabaszcz, Thomas, & Mendoza, 2014).

Clinical Trial NCT00312962. This randomized clinical trial of cognitive training for people with schizophrenia was registered at ClinicalTrials.Gov in 2006, although testing had begun 2 years earlier. The study was completed in 2013. The registration lists enrollment of 80 participants in total; participants in matched pairs were randomly assigned to a targeted cognitive-training group or a computer-games group and completed 90 hours of training over 20 weeks. However, none of the papers associated with this clinical trial number reported that many participants, and most reported 50 hours of training rather than 90. The primary outcome measure was described as a neuropsychological battery, with assessments at 8 weeks and 14 weeks as well as a 6-month post-intervention follow-up. Subsets of participants also underwent neuroimaging. The registration itself is incomplete and does not fully document the details of the study. For example, it includes no details about the intervention tasks, outcome measures, or planned analyses. Without such preregistered specifics, it is unclear which analyses were planned and which were exploratory.

By examining the 11 papers from this trial cited by Cognitive Training Data, we were able to determine that the cognitive-training intervention consisted of adaptive auditory training exercises from Posit Science and that the control group rotated through 16 commercially available games (e.g., puzzle games, mystery games, pinball games). Although all participants in this set of papers received 50 hours of auditory training using Posit Science software, subsets of the complete sample apparently received additional training of different types, and analyses of these subsets are reported separately. For example, one subset of participants received an additional 30 hours of visual-processing training and 20 more hours of cognitive-control training (e.g., Fisher, Holland, Subramaniam, & Vinogradov, 2010).

Another set of studies, reportedly run "in parallel" with the larger clinical trial, added 5 to 15 minutes of social-cognitive training to each 60-minute session of auditory training (Hooker et al., 2013; Hooker et al., 2012; Sacks et al., 2013). Although neither paper by Hooker and colleagues described how these participants were connected to the larger clinical trial, Sacks et al. (2013) reported that participants were drawn from the control group in the main trial; after they had completed the 6-month follow-up to that study, they were given 50 hours of auditory training plus 12 hours of social training. It is unclear whether the patients described by Sacks and colleagues are the same ones described in the papers by Hooker and colleagues. Sacks et al. (2013) was described as a pilot study, tested 19 participants, and did not include a control group. Consequently, whether or not it overlapped with the Hooker et al. (2013, 2012) samples, the results cannot provide compelling evidence for gains in performance.

Both papers by Hooker et al. (2013, 2012) compared results from 11 intervention participants to results from 11 control participants. The only difference between them appears to be in the analysis of different brain regions from the same fMRI neuroimaging sessions. The 2013 paper acknowledged that the demographics were the same as those reported in 2012 but otherwise did not acknowledge the overlap in the behavioral data. Both papers reported the same Time × Condition interaction for the emotional-intelligence outcome measure (the Mayer-Salovey-Caruso Emotional Intelligence Test). That measure, the only behavioral measure to differ significantly between the treatment and control groups, was no longer significant when controlling for the age difference between the groups (Hooker et al., 2012). Hooker et al. (2012) also found a significant Time × Condition interaction for positive-emotion recognition during the fMRI session but did not observe the predicted interaction for negative emotions. The paper reported "no significant intervention-related improvements in general cognition or functional outcome" (p. 53).

Although the neuroimaging outcome measures reported in these two papers are beyond the scope of this review, the same issue of multiple comparisons arises for these measures as for behavioral measures: Reporting the results for different brain regions in separate papers without correcting for multiple comparisons inflates the rate of false-positive results. In the extreme, researchers could examine every brain region and write separate reports for those revealing statistically significant results. Doing so without correcting for multiple comparisons would inevitably introduce false results in the literature. Collectively, these three papers showed no evidence for differential improvements on the same behavioral outcome measures as the main clinical trial, and they show no benefits of the social-cognitive training on the additional social outcome measure.

Two additional papers associated with Clinical Trial NCT00312962 (Subramaniam et al., 2012; Subramaniam et al., 2014) included 50 hours of auditory training followed by 30 hours of visual training and 10 hours of the same social training described by Hooker et al. (2013, 2012) and Sacks et al. (2013): 15 minutes of social training added to each session of visual training. These two studies appear to have tested the same participants using the same training and testing sessions. Yet neither paper explicitly noted whether these outcome measures were collected in the same session with the same participants, and neither explicitly listed all of the outcome measures that were collected during the MRI sessions. It also is unclear which of the participants whose data were reported in other papers later received the additional training described in Subramaniam et al. (2012; Subramaniam et al., 2014) or whether any of these participants were the same as those reported by Hooker et al. (2013, 2012) or Sacks et al. (2013).

Subramaniam et al. (2012) reported the effects of the cognitive intervention on source-memory performance (reality monitoring), with testing occurring in the scanner before and after training. Subramaniam et al. (2014) reported the results of 1- and 2-back working-memory tasks along with fMRI activations corresponding to those tasks. Training yielded improvements in both the 1-back and the 2-back tasks. The reported Time × Condition interaction included a healthy control group in addition to the matched training and control groups of participants with schizophrenia, meaning that it was not a focused test of the differential benefits of training. Follow-up tests did reveal significantly greater improvements for the intervention group than the matched control group for the 2-back task. Note, though, that this differential improvement partly reflects an unexpected decrease in performance in the second session for the control group.

To the extent that the pattern of findings across these studies is robust, something that is hard to determine given the degree of overlap, they could constitute evidence for some degree of transfer of training. Although the trained tasks did involve learning and memory, they did not specifically train source monitoring or *n*-back. Whether these findings reflect narrow or broad transfer depends on whether the training and outcome tasks can be considered to tap the same underlying cognitive mechanism. Still, without correction for multiple comparisons and full reporting of outcome measures, these results should be treated as preliminary and in need of independent replication.

Several other papers associated with this clinical trial reported non-behavioral measures and correlated those with training effects, testing the same set of 55 patients reported in Fisher, Holland, Merzenich, and Vinogradov (2009) and Adcock et al. (2009). These papers reported biomarkers of training effects but presented no new behavioral evidence for transfer of training (Vinogradov, Fisher, Holland, et al., 2009; Vinogradov, Fisher, Warm, et al., 2009).

Collectively, the behavioral data from Clinical Trial NCT00312962 show some improvements in a composite global-cognition measure from 50 hours of intensive auditory perception and memory training (sometimes combined with 30 hours of visual and social training). The benefits to the global-cognition measure were mostly from relative improvements in measures of verbal learning and memory. These studies showed no substantial evidence for improvements in processing speed, cognitive control, visuospatial processing, problem-solving, or executive control. Moreover, behavioral data from a smaller sample of participants (possibly overlapping with the main sample) who underwent both auditory training and social-cognitive training revealed no significant cognitive improvements relative to the same control condition. More broadly, none of the papers reported differential improvements on a quality-of-life measure, and none of the studies included corrections for multiple comparisons either within or across papers.

The tasks that did show differential improvements following cognitive training were distinct from the tasks that were used during training, but they tapped some of the same underlying constructs (e.g., learning and memory). Improvements were limited to these trained domains, suggesting relatively narrow and focused training benefits rather than broad improvements to cognition more generally. The robustness of these effects is unclear, given the lack of complete reporting and direct tests of the critical hypotheses.

Clinical Trial NCT00694889. One additional article from a different clinical trial came from some of the same authors involved in Clinical Trial NCT00312962 (Fisher et al., 2015). This trial adopted essentially the same design as Clinical Trial NCT00312962, started in August 2007, and completed data collection in February 2015. Unlike the earlier study, this one focused on recent-onset schizophrenia in 144 patients aged 12 to 35. It used a comparable battery of cognitive measures, and the training consisted of 40 hours of Posit Science auditory and visual training, compared to a control condition that consisted of playing commercially available games. The registry also noted that participants underwent blood tests, EEG, and MRI.

Fisher et al. (2015) compared improvements in the symptoms and cognitive performance of 43 people who received Posit Science auditory training and 43 people who played commercial video games. The trained participants showed greater improvements on measures of global cognition, verbal memory, and problem-solving (when controlling for age, testing site, and hours of training—the uncorrected results were not provided). The global-cognition measure was just the average of the *z* scores for all other measures, so it included the verbal-memory measure (the tests were not independent). Moreover, the analyses did not correct for multiple comparisons. After correction for those (at least) 12 primary outcome measures that were reported in the paper, the problem-solving measure would not be significant. Note also that the effect for verbal memory depended in part on an unexpected decline in performance in the control group. That pattern somewhat undermines the conclusion that the difference between conditions resulted from training benefits. None of the other measures showed a statistically significant difference in improvement between the trained and control participants, including the two measures of global functioning. Consistent with the behavioral outcomes from Clinical Trial NCT00312962, the first paper based on this clinical trial suggested a possible benefit of 40 hours of auditory perception and memory training on a verbal-memory task, an example of near transfer to an untrained task. It showed relatively little evidence for transfer beyond another laboratory measure of auditory memory.

A cognitive intervention for schizophrenia (three papers). Three papers reported results from the same cognitive intervention for adults with schizophrenia (Kurtz, Seltzer, Fujimoto, Shagan, & Wexler, 2009; Kurtz, Seltzer, Shagan, Thime, & Wexler, 2007; Kurtz, Wexler, Fujimoto, Shagan, & Seltzer, 2008). The study compared the cognitive improvements of 23 participants who received 100 hours of training on an array of 13 cognitive tasks (tapping speed, attention, and memory) that increased in difficulty over the course of 12 months of training against the performance of a control group of 19 participants who received 100 hours of computer-skills training (on Microsoft Office programs). Kurtz et al. (2007) reported the main behavioral outcomes for this study. Kurtz et al. (2008) focused on predicting differences in real-world functional status after the intervention from pretest scores, but the analyses collapsed across the training and control groups, meaning that the results cannot provide additional evidence for training benefits. The 2008 paper provided additional details about the intervention that were not mentioned in the 2007 paper, including that data were collected between 2001 and 2007, that patients were drawn from several sources, and that participants received other social rehabilitation during the period of the study. More importantly, it identified several other measures that were not described in the 2007 paper (e.g., the UCSD Performance-Based Skills Assessment and the Penn Continuous Performance Test). It is possible that these measures were collected only before or after training and not at both time points. Kurtz et al. (2009) examined learning in the treatment group without comparing its performance with that of the control group. This paper did report the addition of 17 more intervention participants, though, implying that the 2007 paper might not have reported the final results from the complete intervention study.

Kurtz et al. (2007) grouped the 16 outcome measures into five cognitive composites (working memory, verbal episodic memory, processing speed, visual episodic memory, and reasoning and problem-solving) that each was tested in a 2 (treatment/control) × 2 (pre- and post-test) analysis of variance (ANOVA). Although clustering the individual outcome measures into a smaller number of composites reduced the number of statistical tests, the paper did not correct for having conducted five tests. Moreover, significant interactions were then followed up with tests of the individual tasks comprising that composite. Of the outcome composites, only working memory showed a significant difference in improvement between the training and control groups, and that effect would not have been significant had the analysis controlled for multiple comparisons. Of the three tests that contributed to the working-memory composite, only the digit-span measure showed a differential improvement between the training group and the control group. That effect also would not have been statistically significant at $p < .05$ after correcting for the three outcome measures tested in this way. In short, the study provides little evidence that 100 hours of cognitive training produced any differential benefits for the closely related outcome measures and no compelling evidence for transfer of training.

Another cognitive intervention for schizophrenia (two papers). Two papers examined the effectiveness of approximately 20 hours of Posit Science auditory training (n = 20) or COGPACK training (n = 19) on the cognitive performance of people with schizophrenia (Popov et al., 2011; Popov, Rockstroh, Weisz, Elbert, & Miller, 2012). COGPACK is a battery of 64 cognitive tasks that is commonly used for cognitive remediation in schizophrenia therapy. The studies were primarily focused on neuroimaging, but the 2011 paper also reported behavioral outcomes for 35 of the participants, including measures of immediate recall, working memory, delayed recall, and verbal fluency. Although these papers each reported results from the same intervention, they fully acknowledged the extent of overlap and the source of the data. Popov et al. (2011) observed differential improvements in sensory gating, something specifically trained by the auditory intervention. This paper also reported greater improvements following auditory training on immediate recall and working memory, with no differences for the other two measures. Note, though, that these significance tests were not corrected for multiple comparisons, and they would not have been statistically significant at $p < .05$ after Bonferroni correction. Popov et al. (2012) examined whether training effects and task performance were modulated by oscillatory brain activity but did not add new behavioral evidence.

In sum, these papers provide evidence that the Posit Science auditory training led to greater improvements in a measure of auditory sensory processing but provide only weak evidence for transfer to other cognitive measures. That said, the control condition used in this study constituted an intervention in its own right, one that might be expected to yield improved performance on some of the cognitive outcome measures. Thus, the relative lack of differential improvements from training does not necessarily imply the lack of effectiveness of the intervention. Ideally, the trained group would have been compared to a matched active control group that would not be expected to show improved performance on the outcome measures. Still, the findings from this intervention provide little evidence for transfer of training to a broader array of cognitive abilities or to real-world performance.

Auditory training for older adults (two papers). Two papers reported the results of a 40-hour intervention comparing Posit Science auditory training (using Brain Fitness) to viewing educational DVDs; one of these papers focused on immediate benefits (S. Anderson, White-Schwoch, Choi, & Kraus, 2013), and one examined performance in a 6-month follow-up (S. Anderson, White-Schwoch, Choi, & Kraus, 2014). In addition to electrophysiological measures, the primary cognitive outcome measures were the Quick Speech-in-Noise Test, the Memory for Words subtest of the Woodcock-Johnson III (auditory short-term memory), and the Visual Matching subtest of the Woodcock-Johnson III (processing speed). The auditory training group showed greater improvements than the educational-DVD group on all three outcome measures. The speech-in-noise task effectively tests the trained auditory skills. The auditory-short-term-memory outcome task tapped the trained auditory memory abilities with a different task. The training also included elements of processing speed. Consequently, the improvements on all three outcome measures constitute fairly narrow transfer or reflect improvements on the trained skills. S. Anderson et al. (2014) included the pre-testing session, the immediate posttest session, and the 6-month follow-up session in a single ANOVA, making it difficult to evaluate the preservation of differential benefits (the ANOVA result could be significant because of the first two sessions even if differential benefits were lost by the final session). In this analysis, the speed-of-processing effect remained statistically significant, but the verbal-memory and speech-in-noise effects did not. Collectively, these two papers suggest that auditory sensory and memory training might yield benefits for auditory perception and memory tasks and for speed of processing, some of which might be short-lived. Improvements on these tasks constitute relatively narrow transfer, and the paper reported no outcome measures assessing broader transfer to other cognitive measures. The control condition, which consisted of educational DVD viewing, also was suboptimal as a baseline for an adaptive cognitive-training task because it likely did not equate for participant expectations for improvement or motivation to perform well on the outcome measures.

Action-video-game training for young adults (two papers). Two papers compared the relative benefits of 50 hours of training on an action video game versus 50 hours of playing a control game for aspects of visual perception, although neither paper described the overlap in their training groups (Li, Polat, Makous, & Bavelier, 2009; Li, Polat, Scalzo, & Bavelier, 2010). The two papers described the same gaming interventions, durations of training, and time to complete training, and several paragraphs in the method section and appendix in Li et al. (2010) were taken verbatim from the supplementary materials for Li et al. (2009). The two papers reported data from different numbers of participants, so it is unclear whether these represented separate subsamples from a larger study, whether all participants completed both tasks but the papers reported results from different time points in the same intervention, or whether different inclusion/exclusion criteria were used to select the participants reported in the two papers from a larger total sample. Li et al. (2009) reported a differential improvement in contrast sensitivity—an enhanced ability to detect shapes that differed subtly from the background (e.g., a grating composed of different shades of gray against a gray background). Li et al. (2010) reported a differential benefit in the perception of a subtle target that was immediately followed by similar targets (a mask); the performance of those who played the action game was less affected by the mask after training. Neither paper mentioned the existence of other measures or the other paper, but with a 50-hour intervention, it is unlikely that only one or two outcome measures were collected (indeed, the study does appear to have included more outcome measures; see C. S. Green, Sugarman, Medford, Klobusicky, & Bavelier, 2012). Typically, extensive interventions like these include the collection of a large battery of outcome measures before and after testing. If other outcome measures were collected but not identified or reported, all statistical significance tests are suspect.

Individual papers reporting all results from an intervention. The remaining studies with active control groups each appear to have been reported entirely in one paper rather than distributed across multiple papers. Most tested older adults with or without cognitive impairments. Several did not directly compare the improvements between the treatment and control groups, instead analyzing improvements separately in each condition

(Mahncke, Bronstone, & Merzenich, 2006; Mahncke, Connor, et al., 2006; Shatil, 2013). Without testing the Time × Condition interaction, such studies cannot provide compelling evidence for the benefits of an intervention. Another study included an additional unmatched control sample as part of the Time × Condition interaction, making it impossible to directly compare the relative improvement in the appropriately matched training and control groups (e.g., unimpaired participants in the unmatched control group versus impaired drivers in both the experimental and control groups; Roenker, Cissell, Ball, Wadley, & Edwards, 2003).

Several of the cited papers reported little benefit of training. For example, a study comparing a group that used Posit Science auditory training software (the program was not specified) to a control group that did a variety of other activities (listening to audiobooks, reading online newspapers, playing a computer game) showed no significant differences in improvement on a wide variety of outcome measures, even without corrections for multiple comparisons (Barnes et al., 2009). Another study showed a greater improvement for a Posit Science–trained group on one of more than 15 measures after 20 hours of training, but this already small effect was no longer significant after 40 hours of training (Keefe et al., 2012); no other outcome measure showed differential improvements. One study showed improvements on one of three primary measures (selective attention on the UFOV) for a group trained on action video games relative to a no-contact control group, but not relative to a group that played an active control game (Belchior et al., 2013).

Other papers on studies with active control conditions reported significant Time × Condition interactions that were at least partly driven by a decline in performance in the control condition rather than by improvements in the intervention group (Rosen, Sugiura, Kramer, Whitfield-Gabrieli, & Gabrieli, 2011). Whether the worsened control-group performance is consistent with age-related declines over the course of the studies or whether it reflects some form of sampling or measurement variability is unclear. In general, practicing a task by retaking it yields improvements. Interpreting relative benefits when the control group shows declining performance from pretest to posttest becomes difficult, especially if the duration of the intervention is too short for age-related declines to plausibly occur.

Most of the papers for this set of studies showed improvements on the trained tasks, as expected, but little or no differential transfer to other tasks (e.g., Loewenstein, Acevedo, Czaja, & Duara, 2004; Pressler et al., 2011). For example, Pressler et al. (2011) compared heart-failure patients trained with Posit Science Brain Fitness to a group who read health-related information, and tested performance on a large battery of outcome measures. The intervention group improved significantly more on only one of the 12 outcome measures, and that one result would not have survived correction for multiple comparisons.

Those studies that did find differential improvements mostly observed them for tasks closely related to the trained ones. For example, several trained participants on laboratory-based auditory tasks using Posit Science software and found greater improvements (relative to the control condition) on different auditory lab tasks that tapped similar underlying cognitive abilities (S. Anderson et al., 2013). And training on a speed-of-processing task (the UFOV) led to improved performance on the UFOV (Roenker et al., 2003; Vance et al., 2007) or on a measure of choice response time (Roenker et al., 2003).

Several papers used a battery of cognitive tasks for training and then measured performance on a different battery of outcome measures designed to measure the same underlying constructs (e.g., Lampit et al., 2014; Loewenstein et al., 2004; Peretz et al., 2011; Shatil, 2013; Shatil, Mikulecka, Bellotti, & Bureš, 2014). Differential benefits from training typically were limited to a few of the outcome measures, even without corrections for multiple comparisons. These studies provide some limited evidence for near transfer.

One paper compared training on an action video game that demanded multitasking (NeuroRacer) to both training on a single-task version of the game and a no-contact control condition in a group of 46 older adults (Anguera et al., 2013). Of the 11 reported outcome measures, the multitasking group improved significantly more than the single-task group on three (without correction for multiple comparisons). The paper did not report pretest and posttest means, making it unclear whether the groups were equated before training. Although the study included a well-matched, active control group, it tested small samples (15 or 16 participants per group) and provided little compelling evidence for transfer of training. (Note that this paper received extensive media coverage and that the training game is the basis for products being developed by Akili Interactive Labs. See Simons, 2013, for an in-depth post-publication review.)

Taken as a whole, these 16 papers provide relatively little evidence for broad transfer from cognitive training to distinct cognitive outcome measures. Several provide no evidence for training benefits; others do not report the crucial statistical tests necessary to evaluate whether or not training led to differential improvements; and those that did conduct the appropriate statistical tests often did not correct for multiple comparisons. The few studies that did find training benefits generally observed them for a small subset of measures that tapped the trained skills.

Summary of the evidence cited by Cognitive Training Data. The studies cited by Cognitive Training Data provide little compelling evidence for transfer of training. Perhaps the most robust benefits accrue from speed-of-processing training, but those benefits tend to be limited to speeded performance on other laboratory tasks. The one exception might be a trend whereby speed-of-processing training benefits timed IADLs. Ideally, future studies should examine the robustness of training benefits for such measures, because they represent a rare example of transfer of training to objectively measured real-world activities.

For those studies that did use an active control group as a baseline, the activities the group completed were typically not closely matched to the tasks of the intervention groups. Active control groups often completed tasks such as watching educational DVDs, learning how to use Microsoft Office programs, or completing crossword puzzles—poor matches for an intensive, adaptive, and difficult cognitive-training regimen. Although active control groups are preferable to passive ones, without appropriate matching of the demands of the tasks, participants likely will have different expectations for improvements. Consequently, differential improvements might reflect experimenter demand characteristics or a differential placebo effect. None of the cited studies tested whether the intervention and control groups had different expectations for improvement on the tested outcome measures.

The papers cited by Cognitive Training Data suffer from a number of limitations in addition to unmatched control groups. For example, almost none of the papers corrected p values for multiple comparisons, and most appear not to have reported all of the tested outcome measures. Also, many of the cited papers reported evidence that overlapped with that reported by other papers, and that lack of independence means that there are far fewer than 132 independent tests of the effectiveness of cognitive interventions.

Several of the large-scale studies that did show some differential benefits combined cognitive-training interventions with other interventions, such as strategy training or vocational training. The use of such scaffolding might well produce better outcomes, either because those aspects of training are themselves effective or because they work in synergy with cognitive training. Unfortunately, this experimental strategy also makes evaluating the specific benefits of cognitive training more difficult. If improvements could potentially be attributed either to these noncognitive components or to their interaction with cognitive training, it is not clear whether the cognitive component alone is necessary—let alone sufficient—to produce benefits.

The populations of people trained in many cited papers have differed substantially from the target market for brain-training products. Most brain-training products are marketed to aging adults or to children with learning disabilities. Some, like Lumosity, are marketed mainly to the general public. Yet many of the studies cited by Cognitive Training Data focused on patient populations with cognitive deficits, with the largest subset focusing on training for people with schizophrenia. It is not clear whether results of interventions for patient populations suffering from cognitive deficits will generalize to populations experiencing typical cognitive aging or exhibiting typical adult levels of functioning. Results from studies with cognitively impaired elderly samples might not generalize to less impaired populations, and vice versa. Without explicit and direct testing of products with the target demographic, generalization to or from other populations is not justified.

Additional evidence cited by brain-training companies

This section reviews the evidence cited by 12 companies listed by SharpBrains as having a substantial research or market presence in the brain-training domain whose websites also cited published, peer-reviewed journal articles in support of their claims. Here, we focus on those citations not already covered in our review of ACTIVE and Cognitive Training Data. Five of the largest companies provided substantially more citations than the remaining ones, and we focus on them first. We then review the limited number of citations from the remaining seven companies. Note that many of these citations are to foundational papers or to interventions that used tasks not involving the company's own products (e.g., many companies cite studies from ACTIVE).

Scientific Learning Corporation. Formed in 1996 by Michael Merzenich, Paula Tallal, Bill Jenkins, and Steve Miller, Scientific Learning Corporation sells Fast ForWord, a training program that is based on research by the company founders. Fast ForWord provides practice in auditory and visual discrimination, attention, and working memory, and it emphasizes vertical transfer to developing reading skills. The ability to discriminate among phonemes and manipulate them in memory (skills collectively described as "phonological awareness") is central to reading development (Bradley & Bryant, 1983), and there is good evidence that focused practice with phonological awareness can help deficient readers (Bradley & Bryant, 1983; Brem et al., 2010; Temple et al., 2003). In Fast ForWord, children spend up to 100 minutes per day, 5 days per week, for a total of 30 to 100 hours of training that consists of performing tasks that involve listening to speech with modified temporal properties and frequency patterns. The program is targeted mostly to students with language-learning difficulties and can be used either at school or at home. Although Fast ForWord mostly targets reading, some of the claimed benefits are more general. For example, the Scientific Learning Corporation website claimed that Fast ForWord applies "proven research on how the brain learns to accelerate learning [and to] produce patented solutions essential for academic, career, and lifelong success" ("About Us," 2015, para 2).

None of the papers listed by Scientific Learning Corporation were cited by Cognitive Training Data. Of its 51 citations, 25 were not peer-reviewed publications, five were correlational, two were reviews or meta-analyses with no new data, one presented a between-groups comparison without an intervention, one was a technical description of the auditory-processing algorithms used for Fast ForWord, 10 reported interventions that lacked any control group (many of which were case studies of a few children undergoing Fast ForWord training), and six reported interventions with two or more matched groups without random assignment to conditions. Only one paper reported a randomized controlled trial (Gillam et al., 2008). We first evaluate the evidence from studies using matched controls and then consider the one randomized controlled trial.

One of the matched-groups studies focused only on neuroimaging outcome measures, comparing a group of 5 children who completed 50 hours of Fast ForWord training to a 6-child no-contact control group (Russo, Hornickel, Nicol, Zecker, & Kraus, 2010). The analyses adopted a case-study approach, comparing improvements for each trained child to the mean improvement for the no-contact group. In this case, children were assigned to the no-contact group if they did not want to commit the time necessary for the intervention, potentially introducing a substantial self-selection bias. Another study (Stevens, Fanning, Coch, Sanders, & Neville, 2008) compared two groups of children with typically developing language—a group trained on Fast ForWord (n = 9) and a no-contact control group (n = 13)—to a group of children with specific language impairment (n = 7). The language-impaired children showed improvements, but the typically developing children did not. Unfortunately, the analyses combined all three groups into a single Time × Condition ANOVA, making a direct comparison of the two matched groups (language-typical children with and without training) impossible. The lack of a direct comparison of the matched groups undermines any inferences about the effectiveness of training in this study.

One early matched-group study trained 22 typically reading children for 100 hours on tasks that used either modified or normal speech (Tallal et al., 1996). The modified-speech group showed bigger improvements on an auditory-processing task, providing evidence that auditory training could enhance performance on other auditory-processing tasks. A related study (Habib et al., 2002) trained participants by having them perform an auditory odd-one-out task that used either normal (n = 6) or modified speech (n = 6). Unfortunately, the study did not report a statistical test of the difference in improvements across conditions for the primary outcome measure, and only one of a number of secondary measures appeared to show a differential improvement (without correction for multiple comparisons). Subsequent studies reported in the same paper did not include a comparison group for the modified-speech condition.

Another study using a matched-groups design (P. E. Hook, Macaruso, & Jones, 2001) trained children with language impairments on either Fast ForWord or the Orton-Gillingham Language Approach program (a regimen that adopts an individualized, multisensory approach to language training). Groups of approximately 10 children completed each intervention (about 36 and 25 hours of training for the two interventions, respectively) or were part of a no-contact control group. Tests on a large battery of reading and auditory-processing tasks revealed no differential improvements for the Fast ForWord–trained group relative to the no-contact control group. Only one measure differed significantly between the two interventions, and it favored the Orton-Gillingham intervention (although it would not have been significant following correction for multiple comparisons). One other matched-groups study trained 4 children with Fast ForWord and 3 children with Laureate Learning Systems software (Marler, Champlin, & Gillam, 2001). Given the small sample, it is perhaps unsurprising that the study showed no significant differences in improvement.

The only large-scale, peer-reviewed publication cited by Scientific Learning Corporation is also the only randomized controlled trial they cite (Gillam et al., 2008). This study compared groups undergoing three different language-specific interventions (Fast ForWord, Laureate Learning Earobics, and one-on-one work with a speech-language pathologist) to an active control group that played various computer games for academic enrichment that did not specifically target language (e.g., The Magic School Bus). Each group included 54 participants with language impairments who were trained for 50 hours. Outcomes included a comprehensive spoken-language assessment and auditory backward-masking task. The paper did not list the secondary outcome measures, only noting that they would be reported in separate papers. All children in all groups improved on both the language assessments and backward masking. The Time × Condition interaction was not significant for language processing. Only one subtest showed an advantage for the three language-specific intervention groups over the control group, but that advantage occurred only at one time point, and the analysis was not corrected for multiple comparisons. Backward masking similarly showed no differential improvements across groups. In short, Fast ForWord and other language-specific interventions showed no differential benefits for language processing or auditory backward masking.

In summary, the evidence cited by Scientific Learning Corporation provides little compelling evidence for the effectiveness of Fast ForWord as a tool to improve language processing or other aspects of cognition. Studies showing benefits typically included interventions that lacked any control group, and those with a control comparison group generally showed little evidence for differential improvements. The only randomized controlled trial provided no evidence for differential improvements, even on measures tapping similar aspects of auditory language processing.

Posit Science. The Posit Science Corporation was founded in 2002 by Michael Merzenich and Jeffrey Zimman as Neuroscience Solutions Corporation, changing its name to Posit Science in 2005 (Bloomberg, 2016). Merzenich had previously cofounded Scientific Learning Corporation, and he owns the cognitivetrainingdata.org Web domain that hosts the open letter from brain-training proponents.

Posit Science's first brain-training products were Brain Fitness and InSight, both released as DVDs. Brain Fitness included auditory discrimination and attention tasks, and InSight focused on visual processing, memory, and processing speed. In 2008, Posit Science acquired the license for the UFOV training task used in the ACTIVE trial to train speed of processing (Ball et al., 2002) and incorporated the task into its products. In 2012, Posit Science launched BrainHQ as an online training platform that combined some of the tasks from InSight and Brain Fitness. BrainHQ was used in most of the recent Posit Science intervention studies. Many of the claims Posit Science makes about the effectiveness of BrainHQ are based not on interventions using the product itself but on earlier interventions using the UFOV (e.g., ACTIVE). In addition to BrainHQ, Posit Science currently markets Drivesharp (www.drivesharp.com), a UFOV-based training task intended to enhance driving performance.

Of the 107 papers cited by Posit Science at various places on its website, 64 were included in the Cognitive Training Data list. Of the 43 Posit Science citations not listed by Cognitive Training Data, 20 were correlational, nine were reviews or meta-analyses with no new data, seven presented between-groups comparisons without an intervention, one reported a non-cognitive intervention, and one only measured learning during training. Five reported the results of randomized controlled trials. Of those, one reported only health outcome measures and not cognitive ones, and two were from large-scale studies that we already discussed as part of our review of ACTIVE and Cognitive Training Data (one from ACTIVE and one from Clinical Trial NCT00430560). Thus, two of the 43 cited papers provide evidence from randomized controlled trials that we have not yet reviewed (Edwards, Hauser, et al., 2013; Mazer et al., 2003). The lack of additional intervention papers again is to be expected; Cognitive Training Data presumably cited the evidence that Posit Science believes best supports the efficacy of brain training.

Mazer et al. (2003) tested stroke patients who had been referred for driving evaluation, comparing the effectiveness of UFOV training versus training with other computer games (Tetris, Mastermind, Othello, Jigs@w Puzzle) on a large battery of outcome measures and an on-road driving assessment. The UFOV intervention group showed no differential improvements on any of the outcome measures. Edwards, Hauser, et al. (2013) tested Parkinson's patients by comparing a group trained on Posit Science InSight to a no-contact control group. The trained group improved more on the UFOV, a variant of which was used for training. The study found no differences between the groups in measures of depression symptoms or self-reported cognitive functioning. Thus, the two additional randomized controlled trials cited on the Posit Science website provide no support for transfer of training to measures other than those closely related to the trained task.

Lumosity. Lumos Labs was founded in 2005. Unlike those of the other companies described here, its founders were not themselves psychology researchers. Lumos Labs launched its website, Lumosity.com, in 2007. It now has more than 70 million members. Lumosity training involves playing a number of gamified versions of cognitive tasks, such as the Eriksen flanker task and the Corsi block-tapping test. Like Nintendo's earlier Brain Age software, Lumosity gives users feedback about their brain "fitness" and updates that fitness level as performance on the practiced tasks improves. Lumosity is perhaps the most heavily marketed of all of the brain-training products, but relatively little peer-reviewed, published research has examined the effectiveness of Lumosity training for improving other cognitive skills.

The Lumosity website provides a bibliography of 46 citations to research supporting its claims ("The Human Cognition Project: Bibliography," n.d.). Several additional articles come from a compilation titled "Published Literature on Lumosity" that was provided to us by a journalist who had previously received it from a Lumosity employee. (We contacted that Lumosity employee by e-mail to request the file, but did not receive it.)

Of the 55 listed citations, only three were also cited by Cognitive Training Data. Two were from Clinical Trial NCT00312962, which examined Posit Science auditory training as an intervention for cognitive impairments in schizophrenia (Fisher et al., 2009; Fisher et al., 2010), and one was a review article (Fisher, Loewy, Hardy, Schlosser, & Vinogradov, 2013). Of the remaining 52 citations, 36 were not to peer-reviewed journal articles. Most of these were poster presentations at conferences, proceedings papers, or talks; of the remaining papers, one was correlational, one was a review article, two reported between-groups comparisons, two used Lumosity performance as an outcome measure for an exercise intervention, one measured learning only, and four reported interventions with no control group.

The 55 citations yielded five additional papers with randomized controlled trials. Two of those papers (Ballesteros et al., 2014; Mayas, Parmentier, Andres, & Ballesteros, 2014) reported data from the same intervention study in which the intervention group played Lumosity games for 20 hours and the control group occasionally had coffee with the experimenters. This "active" control group partially equated social contact and retest effects across the groups, but it did not equate any other task-based factors that might contribute to differential improvements. Of the two outcome measures reported to be significant by Mayas et al. (2014), one had a *p* value that was rounded down to .05 and the other might not have been significant after correction for multiple comparisons. Follow-up tests did not report the Time × Condition interaction. The study was registered at ClinicalTrials.Gov after this paper was submitted for publication, and that registration mentions several other outcome measures that were not reported in the paper.

Ballesteros et al. (2014) reported the results for a larger battery of outcome measures from the same study (see Ballesteros et al., 2015, for a published correction acknowledging the overlap across papers). In addition to the oddball-task measures reported in both papers, Ballesteros et al. claimed differential improvements in choice response time, immediate and delayed memory for family pictures, and two of five self-reported aspects of well-being. However, the Time × Condition interaction for response time was not significant, and the two working-memory measures were incorrectly treated as two instances of the same measure. Taken collectively, these two papers suggest some possible improvements on measures of memory, response time, and attention, but it is not clear whether any of these outcome measures would have been statistically significant with more appropriate statistical analyses.

M. Finn and McDonald (2011) tested a total of 16 participants, 8 in a group that played 30 hours of Lumosity games and 8 in a no-contact control group. The study included a large battery of outcome measures, and only one of those improved significantly more in the training group than in the control group (with no correction for multiple comparisons). Moreover, the difference appears to have been driven mostly by a decline in performance from pretest to posttest in the control group rather than by an improvement in the trained group. Given the small sample sizes in this study, it lacked statistical power to find small effects. Consequently, it is unsurprising that it did not find evidence for transfer of training.

J. L. Hardy, Drescher, Sarkar, Kellett, and Scanlon (2011) trained 14 participants on Lumosity tasks for approximately 10 hours and compared their improvements on outcome measures to those of 8 participants in a no-contact control group. The paper reported using a battery of outcome measures but did not identify them, and it did not report any statistical support for claims of differential improvement. Consequently, this paper cannot be taken as supportive evidence for transfer of training.

Kesler et al. (2013) trained 21 breast cancer survivors on Lumosity tasks for 20 hours and compared their improvement on a battery of cognitive tasks to that of 20 participants in a no-contact control group. Trained participants showed bigger improvements on the Wisconsin Card Sorting Test, a measure of executive functioning. The authors also found differential improvements on a letter-fluency task, a symbol-search task, and a self-report measure of executive functioning. They noted that a corrected alpha level was used, and the self-report measures were not significant with that correction. Note also that all analyses included age, education, radiation, hormone therapy, clinical depression scores, and time since chemotherapy as covariates, and the results without these covariates were not reported. Given the relatively small sample sizes, it is possible that the inclusion of covariates (or major differences between conditions on these variables) might have altered the reported effects. In sum, this study provides some support for the idea that training improves performance on a handful of cognitive tasks, but it is unclear how broadly the training transferred—the tasks chosen for training were those that targeted executive functioning.

One large online randomized controlled study of Lumosity training appeared after the cutoff date for our selection of papers (J. L. Hardy et al., 2015). The study is notable for its large sample: An initial sample of 9,919 participants were randomly assigned to a Lumosity training group or to an active control group (which completed crossword puzzles), and data from 4,715 were included in the analyses. Both groups were instructed to practice their assigned activities for at least 15 minutes, 5 times per week, for 10 weeks. The outcome measures completed before and after training consisted of a seven-task neuropsychological assessment battery (forward span, backward span, Raven's Progressive Matrices, grammatical reasoning, arithmetic reasoning, go/no-go, and a search task) and a self-report assessment of cognition and well-being. Aggregate measures derived from this battery revealed significantly greater improvements for the Lumosity group relative to the control group in both cognitive functioning (Cohen's $d = 0.26$) and self-reported cognition and well-being (Cohen's $d = 0.25$).

These significant effects should be interpreted with caution, given the nature of the sample. Participants were people who already had free Lumosity accounts but had not yet become paid users, and compensation for participation consisted of a 6-month subscription. Consequently, the sample pre-selected for people who already were interested in Lumosity and presumably thought it might have benefits (they likely were exposed to advertisements touting the benefits of Lumosity training as well as testimonials on the Lumosity site). The promised reward, granted only to those who completed training, further selected for participants who valued Lumosity and who believed in its efficacy enough to remain in the study for 10 weeks. Presumably those who are interested in Lumosity would be more highly motivated to perform well during training in the Lumosity condition than in the crossword-puzzle control condition. Consistent with that concern, the rate of attrition was greater for those in the control group (53%) than for those in the Lumosity group (47%). Participants in the control group might well have expected less improvement than those in the Lumosity group did, and it seems likely that many of these participants realized that they were in a control group. Such differential expectations might well account for different improvements on both the self-report and objective performance measures.

Even if the objectively measured training benefits cannot be attributed to differential expectations, motivation, or attrition, it is unclear that the results constitute far transfer, given that the outcome measures were in several cases closely related to the trained Lumosity games. For example, the Follow that Frog game requires the same sort of sequence memory as the forward- and backward-span tasks used in the outcome battery. Similarly, the Raindrops and Multiplication Storm games require the same sort of speeded math as the arithmetic-reasoning outcome measure.

Finally, the findings from this study should be qualified by the author-acknowledged conflicts of interest. The study was funded by Lumos Labs, and five of the seven authors are employees of the company. Another author is a consultant, and the last is on the company's scientific advisory board. Although such conflicts do not preclude rigorous science, readers must be mindful of the incentives they introduce to report positive results.

In sum, Lumos Labs cites little, if any, compelling evidence from randomized controlled trials that supports the claim that practicing Lumosity tasks yields broad improvements in cognitive abilities. Moreover, the evidence the company does cite mostly consists of non-peer-reviewed studies or studies that could not (by design) provide such evidence. The recent large-scale clinical trial is a positive development, but it lacks controls for possible placebo and expectation effects and used an inadequately matched baseline condition.

CogniFit. Founded in 1999 by Shlomo Breznitz, a former psychology professor at the University of Haifa, CogniFit initially focused on cognitive training for driving performance (DriveFit) and later expanded its product line to form a more complete "brain gym" ("Welcome to the Brain Gym," 2004). Their first CD-based product for broader brain training, released in 2004, was called MindFit. That product later was replaced by CogniFit Personal Coach, which in turn was replaced by web-based (2011) and mobile-app-based training (2012).

The CogniFit website cites 29 papers in support of its claims, eight of which were cited by Cognitive Training Data. Of the remaining 21 papers, two were not peer-reviewed journal articles, eight were review articles without new data, two presented correlational evidence, one presented a between-groups comparison, one measured learning but not outcome measures, one was not an intervention study and did not measure cognition, two used exercise interventions rather than cognitive ones, two focused on measuring performance and did not include an intervention, and one compared an intervention group to a separate baseline group that was not tested before and after training (not a true control group). Only one of the papers not also cited by Cognitive Training Data reported the results of a randomized controlled trial with a brain-training intervention (Siberski et al., 2014).

Siberski et al. (2014) used CogniFit software to train 11 adults with intellectual or developmental disabilities for approximately 12 hours. Their performance was compared to that of 11 participants in a video-game control group and 10 in a no-contact control group. Unlike in many intervention papers, the analyses in this paper were corrected appropriately for multiple comparisons. And with that correction, the study found no significant improvements from pretest to posttest and no differences in improvements across groups. Given the small sample, the study was underpowered to detect small effects, and some of the trends might merit a larger follow-up study.

In sum, the citations on the CogniFit website, beyond those also cited by Cognitive Training Data, did not include randomized controlled trials with evidence for transfer of training beyond the trained tasks.

Cogmed. The Cogmed training program was developed commercially from research conducted by Torkel Klingberg at the cognitive neuroscience laboratory at the Karolinska Institutet. His team reported generalizable gains in working memory following a 5-week intervention in children with attention-deficit/hyperactivity disorder (ADHD; Klingberg et al., 2002). Cogmed training focuses on working-memory skills, with the goal of reducing learning and attention problems associated with disorders such as ADHD or memory impairments due to brain injury. Cogmed RM, the most widely used software training program, is designed for school-aged children and consists of a variety of training tasks linked together with a space theme. Other Cogmed products have been developed for younger children (Cogmed JM, which employs only visuospatial activities) and adults (Cogmed QM, which uses modified Cogmed RM activities). The Cogmed company was acquired by Pearson Education in 2010.

Training activities involve recalling increasingly longer sequences as performance improves with practice. The majority of tasks require the retention of visuospatial information (e.g., the order and location of cells illuminated in a grid); other tasks train memory for sequences of digits or letters. Whereas some require only simple serial recall, others require mental transformation (e.g., remembering the relative positions of stimuli on a rotated display or recalling digits in reverse sequence). These additional task demands impose conditions akin to those of complex working-memory span tasks in which participants must remember information while simultaneously processing other information. Trainees complete exercises either at home under remote supervision or in a school or rehabilitation setting for 5 days each week over a period of 5 weeks, totaling 12 to 20 hours of training.

Cogmed standardized its training and, until 2015, made active control interventions available for research use. These controls were non-adaptive versions of the same programs, fixed at a relatively undemanding level (a span length of two items). They provided a useful baseline, mimicking the core task demands in the training group, and also instituted some control for contact with training coaches. Although the use of a standard active control group is laudable, the amounts of time demanded by the adaptive and non-adaptive training regimens were not precisely equivalent (Chacko et al., 2014), and, inevitably, neither were the required effort and engagement. The availability of this standard control program enabled much of the research on Cogmed to adopt broadly comparable designs. Unfortunately, Cogmed's decision to phase out support for this non-adaptive control condition likely will make it more difficult for researchers to use such a well-matched active control condition in the future.

As of July 2015, the Cogmed website listed 46 papers in support of its claims, two of which were cited by Cognitive Training Data. Of the remaining 44 papers, one was not a peer-reviewed journal article, four presented correlational evidence, nine did not include a control group, and three either used some form of matched control group or used trained participants as their own control group. Our review focuses on the remaining 27 papers reporting randomized controlled trials plus one paper reporting a large randomized controlled trial published in March 2016 (Roberts et al., 2016). We included this more recent study because of its unusual size and rigor. Eleven of the randomized controlled trials used an active control group. Most of those used the control condition provided by Cogmed, but a small number used control conditions taxing other cognitive abilities, such as nonverbal reasoning (Bergman Nutley et al., 2011), mathematics (Gray et al., 2012), or inhibitory control (Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009). Fifteen interventions used a passive control group, and two compared the performance of two training groups without considering either a control group.

Most of the randomized controlled trials have included working-memory outcome measures similar in structure to the training tasks but with often-subtle differences in memory items or responses. Training might consist of remembering sequences of dots presented in a 4 × 4 grid on a computer display ("Trained vs. Non-Trained Tasks Cont.", 2015), and an outcome measure might involve the same task but with blocks in the real world and a manual pointing response. The working-memory demands and the structure of the tasks are nearly identical, but the elements presented (dots on a display vs. blocks in the world) and the response modality (clicking with a mouse vs. pointing) differ. We view these as two variants of the same task rather than distinct measures of the underlying latent construct. More remote transfer within working memory would be demonstrated by, for example, recalling the items in a different order (e.g., reverse order if participants were trained to recall them only in forward sequence).

Most sample sizes in Cogmed studies are small. Of the studies with active control groups, only three trained more than 30 people. Of those with passive controls, six included training conditions with 10 or fewer participants. Studies with such small samples lack sufficient power to reliably detect moderate differences in improvement, a factor that is likely to have contributed to variability in the pattern of significance of training outcomes found in studies testing comparable populations. The recent randomized controlled trial (Roberts et al., 2016) tested an unusually large sample of 226 children, aged 6 to 7, with low working memory. The study compared a Cogmed RM–trained group to a passive control group, with outcomes evaluated at 6, 12, and 24 months. The protocol was preregistered (see Roberts et al., 2011).

Many studies also have employed large batteries of outcome measures tapping working memory, executive functioning, and attentional behavior. We summarize the evidence for the effects of training on each of these outcome domains below. As for most brain-training interventions, few of the cited studies used preregistered protocols, identified primary and secondary outcomes prior to data collection, or corrected significance levels for multiple comparisons. Many of the effects reported as significant would not survive such correction.

Transfer to other working memory tasks. Cogmed training benefits are strongest for transfer tasks that tap working memory and impose demands that are highly similar to those of the trained activities. Both Cogmed RM and Cogmed QM provide daily practice in recalling digits in reverse order. Trained participants showed differential improvements, relative to control participants, on measures of backward digit span that differed from Cogmed tasks only in the recall modality (i.e., verbal responding rather than mouse-based selection from a display; Akerlund, Esbjörnsson, Sunnerhagen, & Björkdahl, 2013; Astle, Barnes, Baker, Colclough, & Woolrich, 2015; Brehmer, Westerberg, & Bäckman, 2012; Dunning, Holmes, & Gathercole, 2013; Foy & Mann, 2014; Gray et al., 2012; Gropper, Gotlieb, Kronitz, & Tannock, 2014; Roberts et al., 2016). This pattern holds across studies that used either active or passive control groups. In children with low working memory, the training benefits persisted for 12 months (Dunning et al., 2013). Similarly, multiple Cogmed training tasks require participants to recall sequences of objects in different spatial locations. After training, most studies report significant transfer to other tasks that also require participants to recall (and probably to rehearse) the sequence of spatial locations with only superficial differences in the visual characteristics of the stimuli (e.g., span-board or dot-matrix tests; Akerlund et al., 2013; Astle et al., 2015; Bergman Nutley et al., 2011; Brehmer et al., 2012; Dunning et al., 2013; Foy & Mann, 2014; Gray et al., 2012; Lundqvist, Grundström, Samuelsson, & Rönnberg, 2010; Roberts et al., 2016; Thorell et al., 2009; Westerberg & Klingberg, 2007). In both cases, these improvements might be better characterized as training-task gains than as transfer because the task demands of the training and outcome measures were so closely matched.

Training-related improvements are less consistent when the transfer tasks impose different processing demands, even when the inputs and storage requirements are the same. For example, forward digit span differs from backward digit span only in the order in which the remembered items are recalled. Yet training gains for forward span typically are smaller and less consistent than for backward span, with differential gains relative to both non-adaptive and passive control groups reported in some (Akerlund et al., 2013; Astle et al., 2015; Chacko et al., 2014; Klingberg et al., 2005; Lundqvist et al., 2010) but by no means all studies (Dunning & Holmes, 2014; Dunning et al., 2013; Foy & Mann, 2014; Gray et al., 2012; Roberts et al., 2016). Backward span requires participants to override the more natural and practiced experience of forward recall (e.g., recalling a telephone number). Most participants perform backward recall using successive forward recall, reporting the final item, then the penultimate item, and so on (Anders & Lillyquist, 1971; Thomas, Milner, & Haberlandt, 2003). Given the different demand, practicing this process confers limited benefits at best for the already practiced task of forward recall.

Cogmed training gains are similarly inconsistent for complex working-memory tasks that combine item storage with other processing demands. In listening span, participants make semantic decisions about sentences while retaining the last word in each sentence. Training of adults with brain injuries led to greater improvements on this measure compared to a passive control condition (Lundqvist et al., 2010), but children with ADHD showed no differential improvements when compared to an active control group (Chacko et al., 2014). Children with low working memory improved more than a passive control group on a span task that required them to follow sequences of instructions (Holmes, Gathercole, & Dunning, 2009), although this improvement was not found in a randomized controlled study comparing a similar group of children to an active control group (Dunning et al., 2013). Children with Down syndrome showed no improvements in counting span, another verbal complex span task, after Cogmed JM training (Bennett, Holmes, & Buckley, 2013). A study of children with pediatric cancer reported differential improvements for adaptive training relative to non-adaptive training on a re-sequencing task with letters and numbers, but not with semantic categories (K. K. Hardy, Willard, Allen, & Bonner, 2013). Adults with acquired brain injury did not show the same benefit in letter-number re-sequencing relative to waitlist controls (Akerlund et al., 2013).

On the other hand, evidence for transfer of Cogmed training to other visuospatial complex span tasks is largely positive. On an odd-one-out task requiring memory for the sequence of positions of items that did not match other items, children with Down syndrome showed greater gains after training than a passive control group (Bennett et al., 2013). However, corresponding improvements following training were not observed for 4-year-old typically developing children (Bergman Nutley et al., 2011). Children with low working memory showed selective significant improvements with adaptive compared with non-adaptive training on the Cogmed Mr. X task (which requires visual matching following by retention of spatial locations) immediately after training, but not 12 months later (Dunning et al., 2013). Similarly, training on the same task produced no benefits after 6 months in the recent large-scale clinical trial (Roberts et al., 2016). Finally, children with ADHD showed enhanced performance relative to an adaptive control group in a complex spatial span task (Chacko et al., 2014). Transfer to other spatial-memory tasks might reflect a refined rehearsal strategy (Awh et al., 1999), one rarely practiced in everyday life. Variability in the number and nature of outcome measures, designs, statistical power, and sample characteristics across these studies makes it difficult to draw firm conclusions. However, the bulk of evidence reveals the most consistent transfer when the task demands during training and transfer overlap the most, most notably in backward digit-span tasks and tasks requiring the recall of spatial positions. Gains are less consistent for outcome measures with distinctive properties, such as novel interpolated processing. Together, the evidence points to a gradient of transfer that is governed by the overlap between the trained and untrained tasks (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008; Sprenger et al., 2013; von Bastian, Langer, Jäncke, & Oberauer, 2013).

Some or all of these observed gains might reflect the development and refinement of recoding strategies, although the extent to which strategy changes contribute remains unclear. Practicing strategies for visual imagery and rehearsal can enhance memory-span performance in both children (Johnston, Johnson, & Gray, 1987; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010) and adults (Chase & Ericsson, 1982; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). Post-training interviews of adult Cogmed trainees revealed greater use of grouping strategies following the adaptive than the non-adaptive training (Dunning & Holmes, 2014). If participants develop successful recoding strategies, the gains tend to be tied to stimuli that can be similarly recoded (Ericsson, Chase, & Faloon, 1980). Consequently, such content-specific strategies might contribute to the narrow transfer observed with Cogmed training.

Transfer to other executive functions. Working-memory performance is closely linked to other executive-control functions, including selective attention and inhibitory control (Kane & Engle, 2003), sustained attention (Holmes et al., 2014), and nonverbal reasoning (Kane et al., 2004). Thus, Cogmed training might be expected to enhance performance on tasks that tap these functions.

The possibility that working-memory training might enhance nonverbal reasoning is central to the *n*-back training literature that we address separately below, but it has also been explored in randomized controlled trials with Cogmed training. In the first study of children with ADHD, training led to greater improvements for the trained group than the active control group on the Raven's Progressive Matrices task, although the gains were not significant after a 3-month delay (Klingberg et al., 2005). Subsequent studies have failed to detect differential benefits on Raven's Progressive Matrices or related block-design tasks (Bergman Nutley et al., 2011; Dunning et al., 2013; Thorell et al., 2009; Westerberg & Klingberg, 2007).

Other studies have used the Stroop color-word interference task to examine whether Cogmed training affects the executive functions of cognitive interference and/or inhibitory control. The initial Cogmed study of children with ADHD observed benefits of training on both speed and accuracy of naming the incongruent ink color of color names (Klingberg et al., 2005), a task known to generate response conflict. Whether this improvement reflects enhanced executive control or changes in lower-level processing remains unclear because of the lack of a baseline, neutral control condition. Significant improvements lasting up to 20 weeks after training in a similar Stroop-interference condition (relative to an untrained condition) were reported for a group with acquired brain injury (Lundqvist et al., 2010). Other studies have found no benefit to performance on Stroop-like tasks (Brehmer et al., 2012; Egeland, Aarlien, & Saunes, 2013; Thorell et al., 2009; Westerberg & Klingberg, 2007).

Several studies administered continuous-performance tests that require sustained attention to a task, another executive ability often impaired in children who have ADHD or poor working-memory performance (Holmes et al., 2014). Although one study reported a significant selective benefit of Cogmed training on a sustained-attention task (Thorell et al., 2009), other studies failed to detect selective effects in a Cogmed-trained group relative to either a non-adaptive training group (Chacko et al., 2014; Dunning et al., 2013) or a no-contact control group (Egeland et al., 2013).

The impact of Cogmed training on adults' performance on the Paced Auditory Serial Addition Task (PASAT) has also been explored, with more promising results (Brehmer et al., 2012; Lundqvist et al., 2010; Westerberg & Klingberg, 2007). In this task, single digits are presented at a regular rate and participants must add together the two most recent numbers. Two studies of patients with acquired brain injury reported significant training gains that lasted up to 20 weeks (Lundqvist et al., 2010; Westerberg & Klingberg, 2007). These studies included few participants (*N*s = 10 and 9, respectively) and lacked active control conditions. Brehmer et al. (2012) compared a larger number of trainees (29 young adults, 26 older) to a non-adaptive control group (26 younger, 19 older), reporting a significant Time × Condition interaction that was preserved 3 months later. The consistency of the results for the PASAT across studies is worthy of note and merits further exploration with larger samples.

Transfer to learning. Working-memory performance correlates strongly with academic success in reading, comprehension, and math (Pimperton & Nation, 2012; Swanson & Jerman, 2007; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013) and with norm-based school measures of educational progress (e.g., Gathercole, Pickering, Knight, & Stegmann, 2004). Given these associations, could Cogmed working-memory training improve knowledge and skill acquisition in school? Several randomized controlled trials have measured literacy and numeracy before and after training. In all cases, these studies have reported no selective improvements immediately following training on measures of word reading, sentence comprehension, spelling, or mathematics (Chacko et al., 2014; Dunning et al., 2013; Gray et al., 2012). A randomized controlled trial with children with low working memory reported no selective improvements in either reading or mathematics 12 months after the completion of training (Dunning et al., 2013). Similarly, the recent larger-scale clinical trial with children with low working memory (Roberts et al., 2016) found no training gains in reading, spelling, or mathematics at either the 12- or 24-month follow-up sessions. A limitation of this study was the absence of a pre-training baseline for academic outcomes—those measures were collected only after training, in part because participants in the study began training at the start of their first year of schooling. Thus, there is no evidence to date that Cogmed working-memory training enhances academic performance as measured by standardized tests of literacy and mathematics.

Transfer to behavior. The original randomized controlled trial of Cogmed training for children with ADHD (Klingberg et al., 2005) reported reduced inattentiveness and hyperactivity/impulsivity based on parental ratings. This finding motivated many replication attempts, but several recent meta-analyses found little evidence that Cogmed reduced these symptoms when raters were blind to condition (Cortese et al., 2015; Rapport et al., 2013; Sonuga-Barke et al., 2013). A recent meta-analysis that excluded studies using the Cogmed active control group reported significant reductions in inattentive symptoms both immediately after training and after a 2- to 8-month delay (Spencer-Smith & Klingberg, 2015). However, this more recent meta-analysis incorrectly coded two of the 11 main findings, analyzed only posttest scores without controlling for pretest differences, and did not correct for publication bias in estimating the benefits of training (see comments on the original article: Dovis, van Rentergem, & Huizenga, 2015). These shortcomings substantially weaken the conclusions permitted by the meta-analysis.

One other study in our review set found benefits in parent but not teacher ratings of ADHD symptoms (Beck, Hanson, Puffenberger, Benninger, & Benninger, W. B, 2010), but that study used a passive control condition in which raters were not blind to the condition assignment. Four additional studies showed no differential benefits on ratings of symptoms (Chacko et al., 2014; Egeland et al., 2013; Gray et al., 2012; C. T. Green, Long, et al., 2012). One study of adults with ADHD found reductions in self-reported symptoms among trained participants compared to waitlist controls. Self-reports of this sort are hard to interpret because of demand characteristics (Gropper et al., 2014). More promising findings were obtained from one small-scale study (adaptive training: $n = 12$; non-adaptive training: $n = 14$) providing objective measurements of off-task inattentive behavior in a naturalistic paradigm designed to simulate the attentional demands of a classroom (C. T. Green, Long, et al., 2012). These behaviors showed a selective reduction in children with ADHD who received adaptive training.

A small number of studies have examined the impact of Cogmed training on other daily behaviors (e.g., cognitive failures, depression symptoms, executive-functioning problems). Two studies, one of adults with ADHD (Gropper et al., 2014) and one of younger and older adults (Brehmer et al., 2012), found significant reductions in self-reported everyday cognitive failures following training, with benefits relative to an active control group lasting 3 months (Brehmer et al., 2012). Akerlund et al. (2013) reported a reduction in depressive symptoms among patients with acquired brain injury following Cogmed training, and Lundqvist et al. (2010) reported improvements in personal ratings of occupational performance after training. As both studies used passive control groups, raters were not blind to condition. No significant changes in executive functioning were found following training of children with Down syndrome (Bennett et al., 2013). Finally, one study of pediatric cancer survivors observed improvements in parent ratings of learning problems, but the benefits were not significant in follow-up testing (K. K. Hardy et al., 2013).

This evidence points to limited and sometimes statistically significant beneficial effects of Cogmed training on the attention symptoms of ADHD, most commonly for ratings made either by the trained individual or someone in close contact with him or her and when raters are aware of whether or not the participants received training. Those people closest to the individual might be best able to detect subtle behavior changes. Equally, they might be biased toward perceiving positive effects given their investment in the training (Cortese et al., 2015; Sonuga-Barke et al., 2013). Ideally, more studies in the future will measure behavior objectively (see C. T. Green, Long, et al., 2012). It is possible that even if Cogmed training does not affect behaviors directly, it might yield motivational benefits for adults—that is, people who train with Cogmed might make more of an effort to improve their symptoms—and that motivational effect could have practical value.

Summary. Cogmed is the international market leader in working-memory training and is widely used as therapeutic support for children with problems in attention and learning. The randomized controlled trials provide strong evidence that Cogmed training improves performance on other working-memory tasks with similar processing demands. Training leads to consistent benefits for tasks that are trained directly, such as recalling digits in reverse sequence and remembering the spatial locations of objects in arrays. Improvements for other working-memory tasks such as forward digit span and complex span measures are less robust and appear inconsistently across studies. Transfer gains are largely attributable to the transfer tasks' similarity to the trained tasks, even within the category of working-memory tasks on which improvements are often classified as "near" transfer.

There is little support for the claim that Cogmed training enhances nonverbal reasoning, selective attention, or sustained attention or that it boosts academic learning. Findings that Cogmed training reduces symptoms of ADHD or has other behavioral benefits typically have not been replicated consistently, and the largest reported improvements have come from studies with designs most subject to reporting biases. Collectively, the papers cited by Cogmed in support of the effectiveness of its brain-training software provide limited and inconsistent evidence for improvements beyond the trained tasks.

The designs and methods employed in most Cogmed training studies, with relatively small sample sizes, large batteries of outcome measures, and infrequent correction for multiple comparisons dominating, have likely contributed to inconsistency in the pattern of outcomes across studies. Nonetheless, these studies have generated reliable evidence of transfer to those untrained tasks that overlap highly with trained activities. However, the studies conducted to date do not provide support for generalized training gains for key outcomes such as executive functions, learning, or the inattentive symptoms of ADHD.

Evidence cited by other brain-training companies. Seven additional brain-training companies on the SharpBrains list cited published, peer-reviewed research (typically in addition to citing many unpublished reports and conference presentations). Many of those published papers, however, did not include randomized, controlled trials.

Advanced Brain Technology (www.advancedbrain.com) cites five published papers on studies of the effects of listening to classical music on cognition, but four were case studies of 1 to 3 children and one was a small study with 9 children. None included a control group and none was a randomized, controlled trial.

Akili Interactive Labs (www.brain.akiliinteractive.com) does not cite any publications on its website, but its ongoing interventions build on a game-training study that was also cited by Cognitive Training Data and reviewed in the "Individual Papers Reporting All Results From an Intervention" subsection (Anguera et al., 2013).

Applied Cognitive Engineering (www.intelligym.com) currently focuses on sports-related interventions, and none of its cited papers examined the effects of cognitive training on sports outcomes. In addition to one study from Cognitive Training Data (Ball, Beard, Roenker, Miller, & Griggs, 1988), its website cites three papers. One compared the benefits of 1 hour of flight-simulator training on later actual flight performance, relative to performance by a no-contact control group (Dennis & Harris, 1998). Another examined the effects of 10 hours of Space Fortress training on flight performance in Israeli cadets relative to a separate group of cadets who received no practice. This second study was not a randomized controlled trial, though—the control group was added later. Finally, Applied Cognitive Engineering cites a study of the effects of different forms of training (emphasis on the whole task or emphasis on parts of the task) on dual-task performance (Kramer, Larish, & Strayer, 1995). After three sessions of training, those for whom parts of the task were emphasized during training improved more on the trained task and showed near transfer to another laboratory dual-task measure (although inferring that benefit required an omnibus ANOVA interaction among four factors).

Dakim BrainFitness (www.dakim.com) cites two papers on the Cognitive Training Data list (K. J. Miller et al., 2013; Willis et al., 2006). Their website also cites five correlational/longitudinal studies, one between-groups comparison, and one randomized controlled trial (Knapp et al., 2006). Knapp et al. (2006) compared the effects of approximately 10 hours of cognitive stimulation in a group-therapy setting to a passive control and reported non-cognitive outcome measures only (cost-effectiveness of and use of services). The paper does not make clear what sort of cognitive stimulation participants received, so it is not clear that it constituted brain training. Moreover, none of the individual measures of service use was significant; only when combining across all measures was the Time × Condition interaction significant, and it would not be with correction for multiple comparisons.

Learning Enhancement Corporation (www.mybrainware.com) cites two papers, one of which was a randomized controlled trial (Avtzon, 2012). In that study, 40 children with learning disabilities received 30 hours of training with BrainWare Safari (20 cognitive tasks) or received schooling as usual (passive control). The primary outcome measure was a battery of cognitive tasks (Woodcock-Johnson reading and math batteries). It is unclear how closely related the training tasks were to the outcome measures, but the trained group improved significantly more than the passive control group on almost all outcome measures. In many cases, the differences were equivalent to 1- to 3-year gains relative to normed scores, bringing learning-disabled children up to age-appropriate norms, a remarkable gain for 30 hours of training. A major concern about this study, though, was the substantial age difference between the treatment and control groups. Although children were drawn from the same grade level, those in the experimental group were 10 months older, on average (9.7 years vs. 8.9 years), suggesting that with the small sample, the findings potentially could have resulted from a failure of randomization to account for group differences. Consequently, this finding requires replication with a larger sample.

Neuronix (www.neuronixmedical.com) cites one intervention study that lacked a control group (Bentwich et al., 2011) and one randomized controlled trial (Rabey et al., 2013). Both were small-sample studies examining the effects of a cognitive intervention combined with repetitive transcranial magnetic stimulation (rTMS) on cognitive measures. It is unclear whether the main intervention group in the two papers was the same. Rabey et al. (2013) compared that group to participants who received sham rTMS and who watched nature videos. However, the paper reported no Time × Condition interaction effects or test statistics for any differential improvements.

Scientific Brain Training (www.scientificbraintrainingpro.com) cites seven published papers. One was an observational study, and two examined the effects of training on the trained tasks and did not explore other outcome measures. One small-sample study (total N = 17) of people with hoarding disorder compared the effects 16 hours of training with Scientific Brain Training versus 16 hours of relaxation training on a battery of outcome measures (DiMauro, Genova, Tolin, & Kurtz, 2014). The study reported a significant Time × Condition interaction for attention but not for working memory or executive functioning. The differential gains on the attention measure appear to have been driven mostly by improved response speed for one of the attention tasks, and the results were not corrected for multiple comparisons. One other randomized controlled trial (Bowie et al., 2013) compared cognitive training to a waitlist control for a group of patients with depression. Although the study observed a differential benefit for attention and verbal learning/memory (without multiple-comparison corrections), the cognitive intervention included substantially more than just the brain-training games; it also included group therapy, training of strategic self-monitoring, bridging to daily life, and homework. Consequently, it is not possible to isolate the effects of the cognitive intervention from those of the other aspects of training. Two additional studies compared different types of cognitive interventions for schizophrenia but lacked baseline control groups. Bowie, McGurk, Mausbach, Patterson, and Harvey (2014) compared cognitive training to functional training (or both). Like the Bowie et al. (2013) study of patients with depression, the cognitive training in Bowie et al.'s (2014) study of people with schizophrenia included other forms of training and therapy, so the observed benefits of cognitive training on cognitive outcomes cannot clearly be attributed to the brain-training software. Franck et al. (2013) compared two different cognitive interventions for schizophrenia and found improvements in the cognitive outcome measures for both. However, the study lacked a baseline control group.

Summary of citations from brain-training companies. The additional citations to the scientific literature listed on the websites of brain-training companies added little evidence for the efficacy of brain-training interventions to that provided by the Cognitive Training Data citations. With the exception of Cogmed, these companies did not provide many citations to additional randomized controlled trials. The Cogmed citations provide some evidence for near transfer but little evidence of broad transfer of training, and many of the randomized controlled trials the company cites used small samples. Most of the randomized controlled trials cited by these companies (but not by Cognitive Training Data) compared the intervention group to passive control groups but still showed little evidence for differential improvements.

Across the seven companies with relatively few citations, only one cited study (DiMauro et al., 2014) compared a brain-training intervention group to an active control group, and that study tested a small sample and did not report outcome statistics. The only study to show seemingly robust outcome effects (Avtzon, 2012) compared the performance of a trained experimental group of children to that of a passive baseline group of children who were substantially younger. That study is an outlier in this literature in that it showed significant differential improvements on a broad range of outcome measures, with almost all measures showing benefits of training. Given the unusually broad pattern of significant improvements, together with the small sample and the age difference between the intervention and control groups, the finding should be treated as tentative until it has been replicated with a larger sample (and, ideally, with an active control group).

As for the Cognitive Training Data citations, many of the studies cited by brain-training companies tested patient populations with cognitive deficits rather than typically developing children or adults. The study of cognitive remediation in patient populations is valuable, but it does not necessarily support generalization of any observed effects to the broader population—and most brain-training companies market their products to an audience broader than the patient populations tested in these studies.

Adaptive Dual *n*-Back Memory Training

Cogmed may be the most prominent company touting the benefits of working-memory training for other aspects of cognition, but it is not alone in making that claim. Recent laboratory studies, independent of brain-training companies, have tested the idea that working-memory training might improve "fluid intelligence," or Gf.

Gf is defined as the ability to solve novel problems and adapt to new situations (Cattell, 1943). The gold-standard test of Gf is Raven's Progressive Matrices, a measure of nonverbal reasoning in which test takers view a 3 × 3 grid of patterns with the lower right cell missing and must select which of eight alternative cells best fits the pattern that logically completes the grid. Gf is highly heritable (Plomin, DeFries, McClearn, & McGuffin, 2008) and stable across the life span (Deary, 2012). It also correlates nearly perfectly with general intelligence (*g*; Gustafsson, 1988), which predicts success in work performance, academic achievement, and even mortality (Deary, 2008; Kuncel et al., 2004). If such associations are causal, then improving Gf might enhance real-world outcomes. Gf also correlates strongly with working-memory capacity, especially as measured by complex span tasks, in which people must simultaneously store and process information ("simple" span tasks require participants only to remember items). For example, a study of more than 250 adults observed a correlation of .64 between the latent construct of working-memory capacity and Gf (Kane et al., 2004).

If working memory correlates strongly with Gf, then training working-memory capacity might enhance Gf, bestowing practical benefits well beyond the trained task. The first exciting evidence for this possibility came from a training study in which participants completed either Raven's Progressive Matrices or a similar reasoning task (the Bochum Matrices Test [BOMAT]) before and after training on a working-memory measure called the dual *n*-back (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). In a dual *n*-back task, participants monitor two streams of stimuli (letters and squares) and press a key whenever a stimulus in either stream matches the one that appeared some number (n) back in the sequence. Participants completed 8, 11, 17, or 19 sessions (each about 25 minutes long) of adaptive training in which the difficulty increased as participants improved.

Not surprisingly, dual *n*-back performance improved substantially. Collapsing across the number of training sessions, those who received training also improved more on the measure of Gf than did those in a no-contact control group. Moreover, the more training participants received, the greater was the gain in the measure of Gf. The effectiveness of this short intervention was dramatic because Gf is thought to be largely stable over time. Previous attempts to increase intelligence through training had met with less dramatic success. For example, in the Abecedarian Early Intervention Project, an intensive intervention from infancy to the age of 5 produced a gain of approximately the same size (6 IQ points) as that implied by the results of just under 6 hours of training in the Jaeggi et al. (2008) study.

This result generated tremendous excitement in the cognitive-enhancement field; *Discover* magazine named the finding one of the top 100 discoveries of 2008. But the original study had serious limitations. First, the sample sizes were small (34 subjects distributed across four training groups and 35 subjects distributed across four control groups). Second, the study used a no-contact rather than an active control condition. Third, it used different reasoning tests with different numbers of items and different time limits across the groups receiving different amounts of training, meaning that they are not directly comparable. Fourth, it used just one measure of Gf for each training group, leaving open the possibility that the results were test-specific. Fifth, at least some of the tested groups completed other transfer tasks that were reported in a 2005 dissertation (Jaeggi, 2005) but not in the 2008 paper, meaning that the statistical tests might not have been adequately adjusted for multiple comparisons. Finally, the crucial Time × Condition interaction was consistent with a training benefit only for the groups trained for 17 or 19 sessions (the significance of the interaction for the 19-session group was reported only in the 2005 dissertation); the groups with less training did not show the same pattern relative to the control group (see Redick et al., 2013, for a discussion of these inconsistencies).

A follow-up training study addressed at least some of these limitations (Jaeggi, Buschkuehl, Perrig, & Meier, 2010). Subjects assigned to complete 20 sessions of either single *n*-back (n = 21) or dual *n*-back (n = 25) were compared to a passive control group (n = 43). All groups completed two tests of Gf (Raven's Progressive Matrices and BOMAT) and a test of working-memory capacity (operation span) in pretest and posttest sessions. Both the single *n*-back and dual *n*-back training yielded improvements on Gf measures in the trained group relative to the control group. However, the pattern of improvements was inconsistent across measures; for the BOMAT, the dual *n*-back group did not improve significantly more than the control group (see Redick et al., 2013, for further discussion). Moreover, there was no evidence for transfer to the "nearer" working-memory measure. As in the original Jaeggi et al. (2008) study, the use of a passive control group means that differential expectations could have contributed to any differences between the trained group and the control group.

In another follow-up study (Jaeggi, Buschkuehl, Jonides, & Shah, 2011), children were trained with single *n*-back training or a knowledge/vocabulary-acquisition task. Overall, there was no evidence for differential improvements in Gf across the training groups. When individuals' data were divided into groups based on how much they improved during training on the *n*-back task, those who showed a large training gain showed Gf improvement. However, median-split analyses like this one are not an ideal way to assess transfer of training (Tidwell et al., 2014).

A study designed to address the limitations of previous studies (Redick et al., 2013; see also Shipstead, Redick, & Engle, 2012, for a discussion of these limitations) trained participants on 20 sessions of the same dual *n*-back task used by Jaeggi et al. (2010) and compared their gains on a battery of 17 cognitive-ability tests (including eight tests of Gf) to those of a no-contact group and an active control group. The active control condition consisted of 20 sessions of training on a visual-search task that is only weakly associated with working-memory capacity. Participants completed the cognitive battery after the 10th and 20th sessions of training. At the end of the study, they also answered questions about their perceived improvements in cognition. Although participants thought they had improved in intelligence, they showed no significant transfer on composite measures of verbal and spatial Gf. Another similar study observed no significant pre- to posttest improvements on Raven's Progressive Matrices in the training group and no differential transfer (Chooi & Thompson, 2012).

More recent studies have produced some evidence for limited transfer. For example, a study of older adults found evidence for training-related improvement in composite measures of working memory and visuospatial skills after 10 or 20 sessions of *n*-back training, relative to a passive control condition (Stepankova et al., 2014). And studies of working-memory training tasks other than the *n*-back have shown some limited transfer. For example, participants who completed 20 sessions of training with complex span tasks showed greater transfer to other complex span tasks than did those trained with simple span tasks or visual search, but the training did not transfer to Gf (Harrison et al., 2013).

Since the original 2008 paper, at least 23 studies have tested the hypothesis that dual *n*-back training enhances Gf. Of these, eight compared improvements in a trained group relative to an active control group, nine used only a passive control group, and another six used both an active and a passive control condition. In many cases, the active control group was not well matched to the demands of the training condition. For example, one study used general knowledge, vocabulary, and trivia questions as a control for single and dual *n*-back training (Jaeggi, Buschkuehl, Shah, & Jonides, 2014).

Several recent meta-analyses have grappled with the conflicting evidence that complex span training can enhance Gf (Au et al., 2014; Karbach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2013), but they have drawn conflicting conclusions. A meta-analysis of 23 studies (Melby-Lervåg & Hulme, 2013) examining training-based improvements in nonverbal ability (including tests of Gf) observed larger gains from pretest to posttest for training groups than control groups (d = 0.19). The observed benefits appeared to be short-lived—by about 8 months later (on average), the effect size for the difference between training and control groups was slightly negative and near zero (d = −0.06). Moreover, the benefit immediately after training largely vanished when only studies that compared the training group to an active control group were included.

In a separate meta-analysis of 20 studies (many included in the Melby-Lervåg & Hulme, 2013, analysis as well) with 98 Gf outcome measures and a total sample size of 1,022, pretest-to-posttest gains were larger for training groups than for control groups, a pattern consistent with the possibility of training-related improvement in Gf (Au et al., 2014). In this analysis, the meta-analytic effect size for the differential improvement was roughly equivalent to an IQ gain of 3 to 4 points. Another meta-analysis that focused on cognitive training in older adults found similar evidence for larger pretest-to-posttest Gf gains for training groups than control groups (Karbach & Verhaeghen, 2014). A re-analysis, however, revealed that several studies claiming training-related gains in Gf could not reasonably be interpreted to support that conclusion (Redick, 2015). For example, differential improvement in the training and control groups in one study resulted at least in part from decreased performance in the control condition (Zinke et al., 2014; see also Schweizer, Hampshire, & Dalgleish, 2011).

Although these meta-analyses showed some support for the effects of working memory training on Gf, the effect again was near zero when considering only those studies that compared transfer from *n*-back training to an active control group (Melby-Lervåg & Hulme, 2015; see also Dougherty, Hamovitz, & Tidwell, 2016). The effect sizes for studies comparing a training group to a passive control group might also have been inflated as a result of pretest differences between the treatment and control conditions. The finding of larger effects in studies comparing performance to a passive control group rather than to an active control group cannot unambiguously be explained by the active/passive difference because those sets of studies differed in other ways (e.g., the amounts paid to participants, whether the study was conducted in the United States; Au, Buschkuehl, Duncan, & Jaeggi, 2016). Collectively, these meta-analyses provide little evidence that working-memory training produces lasting benefits on Gf and other measures of cognition, and effects generally are either weaker or not present when the training group is compared to an active control group. That pattern, coupled with evidence that people believe they are improving in intelligence as a result of training, raises the risk that the observed benefits are due to a placebo effect. The lack of evidence for improvements in Gf is consistent with the literature on the relative stability of Gf across the life span (e.g., Deary, 2012).

The most recent meta-analysis to address the effects of working-memory training on other cognitive measures included not just studies using *n*-back measures but all working-memory studies that compared pre- and posttest scores between an intervention group and a control group (Melby-Lervåg, Redick, & Hulme, 2016). The analysis considered 145 experimental comparisons drawn from 87 publications (Melby-Lervåg et al., 2016). Immediately after working-memory training, participants showed improved performance on closely related working-memory measures but little transfer to other cognitive measures (nonverbal abilities, verbal abilities, decoding, reading comprehension, and arithmetic). On only one of five measures did a trained group show significant differential improvement in comparison to an active control group, and on only two of the five did a trained group show significant differential improvements in comparison to a passive control group. And in no case was evidence for far transfer clear-cut. For example, although the average effect size was significantly different from zero for transfer to nonverbal abilities (when compared to a passive control condition), the studies producing the five largest effect sizes used only a single transfer measure, raising concerns that the observed effects might be tied more to task strategies than to general abilities. And, in several cases, the apparent benefit of the intervention resulted at least as much from a decline from pre-test to post-test in the control group, something we would not expect for relatively stable cognitive-ability measures. That pattern suggests variability in the measurement of the cognitive abilities, meaning that the reported results might have capitalized on noise. Consistent with that interpretation, after correcting for publication bias, studies comparing the performance of a trained group to that of an active control group provided no evidence for differential improvements. More broadly, Melby-Lervåg et al. (2016) concluded that "working-memory training programs appear to produce short-term, specific training effects that do not generalize to measures 'real-world' cognitive skills" (p. 512).

Video-Game Interventions

Shortly after the appearance of the first commercially available video games, scientists began exploring whether playing games might enhance abilities outside of the game itself. For example, not long after the release of Pong in 1972, researchers investigated whether playing it might aid recovery from stroke (Cogan, Madey, Kaufman, Holmlund, & Bach-y-Rita, 1977). Throughout the 1980s, studies reported an association between video-game play and superior perceptual and cognitive skill (e.g., Clark, Lanphear, & Riddick, 1987; Dorval & Pepin, 1986; Gagnon, 1985; Griffith, Voloschin, Gibb, & Bailey, 1983). Many of these early studies just compared gamers to non-gamers, an approach that cannot provide evidence for a causal link between games and cognition; differences between gamers and non-gamers could result from pre-existing differences (i.e., people who possess the superior perceptual and cognitive skills necessary to succeed in these games might be more likely to play them).

A few early studies, however, were interventions that randomly assigned participants to practice or not practice video games. For example, older adults (57–83 years of age) who practiced playing Pac-Man and Donkey Kong improved more in reaction time than did those in a passive control group (Clark et al., 1987). And, after just 5 hours of practice with a fast-paced action game (Robotron), participants showed bigger improvements in some aspects of divided attention than a passive control group (Greenfield, DeWinstanley, Kilpatrick, & Kaye, 1994). Although these early studies were influential, they will not be the focus of our review. First, they typically used designs that permit only weak inferences or no inferences about the causal effects of games (i.e., cross-sectional comparisons or interventions with passive baseline conditions). Second, the games used in these studies are unlike those available commercially today. Video-game research has evolved along with the medium, and most game research today focuses on fast-paced, first-person-shooter games that bear little resemblance to those studied prior to the 2000s.

A paper published in *Nature* in 2003 reinvigorated research into the potential of modern video games to improve perception and cognition (C. S. Green & Bavelier, 2003). The paper reported cross-sectional studies showing that gamers outperformed non-gamers on measures of vision and attention, including a variant of the UFOV, a flanker task, an attentional blink task, and an enumeration task. More importantly, it also reported a training study: A small group of non-gamers (college-aged individuals who had spent little or no time playing video games in the past 6 months) spent 10 hours playing either an action game (n = 9) or Tetris (n = 8), a video game intended to control for improved visual-motor coordination without requiring attention to multiple targets. The action-game group improved more on three of the measures previously shown to differ between experts and novices (UFOV, attentional blink, and enumeration; no flanker data were reported).

This study was notable for (a) supplementing a cross-sectional study with an intervention to explore the causal role of game play, (b) using an active control group, (c) testing an outcome measure (UFOV) known to be associated with real-world performance (automobile crash rates in elderly drivers; e.g., Owsley et al., 1998), and (d) finding improvements with as little as 10 hours of training. It established the standard design for subsequent video-game interventions: Compare improvements on basic measures of perception and cognition following practice with an action video game to improvements in an active control group that played a different game. The study inspired research into the benefits of action games for vision, attention, and executive control.

The effects of action-game training on vision and attention

Initially, action-video-game interventions focused on whether playing 10 or 30 hours of an action game would improve aspects of vision and visual attention more than playing a non-action game (Tetris) for the same amount of time (C. S. Green & Bavelier, 2003, 2006a, 2006b). For example, non-gaming participants randomly assigned to play 30 hours of a first-person-shooter game (Unreal Tournament 2004) showed improvements in their ability to simultaneously keep track of multiple moving objects (multiple-object tracking) from pretest to posttest, whereas Tetris-trained participants did not (C. S. Green & Bavelier, 2006b). Another intervention reported in the same paper found bigger improvements from playing the first-person-shooter game Medal of Honor for 10 hours (again compared to Tetris) on the ability to quickly recognize the number of objects presented on screen (enumeration).

These initial studies reported benefits of action-video-game training on the UFOV, multiple-object tracking, enumeration, and attentional blink. Subsequent intervention papers (sometimes using a strategy game called The Sims as a control in place of Tetris) have reported greater benefits for the action-game group in visual acuity, contrast sensitivity, resistance to visual interference (masking), and perceptual learning after 30 or 50 hours of game play (Bejjanki et al., 2014; C. S. Green & Bavelier, 2007; C. S. Green, Pouget, & Bavelier, 2010; Li et al., 2009; Li et al., 2010).

Although this collection of papers appears to provide extensive evidence for the benefits of action-game training, many of these papers do not provide independent evidence for gaming effects; different outcome measures from the same or partially overlapping samples are reported in separate papers. For example, the same intervention showing an effect of playing Medal of Honor on enumeration was partially reported in two papers (C. S. Green & Bavelier, 2003, 2006b). In that case, the papers acknowledged the lack of independence. In other cases, the existence of overlap between samples reported in different papers has not been acknowledged, leaving readers to infer it from identical game scores across papers (compare the score-improvement functions in C. S. Green et al., 2010, to those in Li et al., 2010; also see scores reported by C. S. Green & Bavelier, 2006a, and C. S. Green & Bavelier, 2007). This lack of independence makes accurate meta-analysis almost impossible, and without a clear and complete description of the nature of the non-independence, it is unclear how much evidence these studies provide for gaming benefits. Moreover, if outcome measures are distributed across papers and not all outcome measures are reported, it is unclear how many other outcome measures did not show any gaming benefits.

Some independent evidence for the potential benefits of action-game interventions come from three additional studies that examined the effect of game training on measures of vision and attention (Feng, Spence, & Pratt, 2007; Wu et al., 2012; Wu & Spence, 2013). In a fairly close replication of C. S. Green and Bavelier (2003), Feng et al. (2007) reported a training benefit on UFOV performance for the action game Medal of Honor relative to the control puzzle game Ballance. However, the samples in this experiment were small (10 participants in each group), and the paper did not report a Time × Condition interaction for the UFOV outcome measure. A separate paper reported EEG components with the same training and control conditions and the same UFOV outcome measure (Wu et al., 2012). Again, the paper did not report a significant Time × Condition interaction, but enough information was reported to calculate approximate Time × Condition p values of .29 and .49 for the two conditions (which varied in the visual eccentricity of the stimuli). Thus, this study did not replicate the original effect (C. S. Green & Bavelier, 2003).

A third paper (Wu & Spence, 2013) compared 10 hours of training on Medal of Honor, Ballance, or a racing game called Need for Speed. Participants were tested before and after training on a variety of visual-search tasks, and differential benefits were observed on response speed in the search task (but not search efficiency) for participants who played the fast-paced games (Medal of Honor, Need for Speed) compared to the control game. Some of the response-speed improvement might have resulted from regression to the mean, because the pretest responses for the Ballance (control) group were substantially faster than for the other groups in the conjunction-search task, and all groups responded with roughly the same speed at posttest (see Wu & Spence, 2013, Fig. 6). In a dual-task search task, search speed did not change, but the ability to accurately report a peripheral target while simultaneously completing the search task did improve differentially for the fast-paced-game groups relative to the puzzle group. These results provide some independent evidence for action-video-game benefits on vision and attention tasks, but the strength of this evidence is unclear.

The effects of action-game training on executive control

In addition to the reported benefits for visual abilities, action-video-game interventions have tested the ability to juggle multiple goals at once in laboratory task-switching paradigms thought to tap executive control. In one study (Strobach, Frensch, & Schubert, 2012), participants were randomly assigned to a first-person-shooter condition (n = 10), a Tetris condition (n = 10), or an additional no-contact control condition (n = 12). After 15 hours of gaming over a 4-week period, participants in the action-game group improved more on measures of task switching and psychological refractory period. Both tasks required participants to keep multiple goals in mind at once and rapidly shift between them in order to make quick and accurate responses.

In a similar study, participants assigned to play 50 hours of action games (Unreal Tournament 2004 and Call of Duty) showed greater improvements in a task-switching task than did participants assigned to play 50 hours of the non-action game The Sims (C. S. Green, Sugarman, et al., 2012). However, after controlling for speed increases across groups, the evidence for improved task-switching ability per se was limited; performance gains were more consistent with faster responses as a result of action-game play than with improved cognitive control.

As for the studies on the benefits of gaming on perception, one of these papers reported data that do not constitute independent evidence. C. S. Green, Sugarman, et al. (2012) noted,

> Subjects completed two experimental blocks . . . as well as several other tasks unrelated to the current paper (e.g., motion discrimination, visual search, contrast detection—however, note: the data presented here was acquired over the course of three separate training studies—and thus the unrelated tasks are not identical in all subjects).

In other words, the paper combined participants from several training studies that presumably tested a variety of different outcome measures. In some ways, combining participants from multiple interventions can be more problematic than reporting outcome measures from a single intervention in multiple papers. First, participants completed different sets of outcome measures, and control participants and intervention participants may not have completed an identical battery of outcome measures (if the control participants came from one of the studies, then their experiences with the outcome measures necessarily differed from those of some of the intervention participants in other studies). Any differences in the nature or content of the cognitive testing before or after the intervention could contribute to differences between the intervention and control groups. For example, if the critical dual-task measure came at the end of a long day of testing for the control participants but at the beginning for some of the intervention participants, that could contribute to an apparent difference in improvements. For a comparison between the intervention group and a control group to be meaningful, all participants in both groups should undergo identical testing before and after the intervention. Otherwise, differences in the nature of the testing might explain any differences that otherwise could be attributed to the intervention.

The effects of action-game training on other cognitive abilities

Although most action-game interventions have focused on attention and executive-control outcome measures, some studies have explored the effects of games on other abilities. For example, participants who played 30 hours of Call of Duty demonstrated increased visual short-term memory capacity and precision relative to participants who played The Sims (Blacker, Curby, Klobusicky, & Chein, 2014). Unlike almost all other brain-training interventions, this study laudably included an attempt to address expectation differences between game groups. However, the effects for memory precision were observed only in one of three conditions, and in that case, the critical interaction of time, game type, and condition (in this case, set size) was marginally significant (p = .056). The critical interaction for the memory-capacity measure was significant (p = .031), but there was no correction for multiple comparisons across the three visual short-term memory outcome measures or any of the other outcome measures described in the supplement.

A study already discussed in our review of attention effects in "The Effects of Action-Game Training on Vision and Attention" above (Feng et al., 2007) was designed as a close replication of the original C. S. Green and Bavelier (2003) study. In addition to measuring the UFOV, the study compared the effects of 10 hours of playing Medal of Honor or Ballance on a mental-rotation task. The results section for the mental-rotation task appears to report a significant Time × Condition interaction. The paper also reported a sex difference in the improvements in each outcome measure in the intervention group, but that analysis was based on a total of 10 participants (3 men and 7 women).

Failures to replicate action-game benefits

The majority of published studies have reported positive effects of action-game play on various perceptual and cognitive abilities, although a few results have called into question the robustness of those effects, and a recent meta-analysis found that the effects of gaming on perceptual and cognitive abilities might be overestimated as a result of publication bias (Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013). One study used a design comparable to that of the original C. S. Green and Bavelier (2003) study but tested more than twice as many participants and trained them for 20 hours rather than 10 hours (Boot, Kramer, Simons, Fabiani, & Gratton, 2008). Despite the larger sample and longer intervention, the action-game group did not improve significantly more than the control group on any of a number of measures of attention and perception, including variants of the UFOV, attentional blink, enumeration, task switching, and mental rotation. Similarly, two studies (one with 10 and one with 20 hours of training) found no differential improvements in speed of processing for an action-game training group relative to non-action-game training group or a passive control group (van Ravenzwaaij, Boekel, Forstmann, Ratcliff, & Wagenmakers, 2014). A study with older adults found no differential improvements on the UFOV between playing Medal of Honor and Tetris (Belchior et al., 2013). The authors labeled Tetris a "placebo" control but subsequently argued that both games might have yielded training benefits in this population. Finally, in addition to the fact that comparisons of gamers to non-gamers should be interpreted with caution, not all of those differences replicate consistently. In a study with more than 800 participants, action gamers did not differ from non-gamers on a variety of measures of attention, reasoning ability, and processing speed (Unsworth et al., 2015). These failures to observe differential benefits of action-game training, coupled with possible publication bias in the field and the lack of independence among papers reporting benefits, suggest either (a) that action-game effects may be smaller than originally reported or (b) that there may be important moderators that determine whether action-game effects are observed.

Training benefits from spatial video games

In addition to action games, a variety of other types of games have been used in intervention studies, including spatial games such as Tetris, real-time strategy games, casual games, and brain-fitness games marketed by video-game companies (e.g., Nintendo). Prior to the recent surge of interest in the effects of action video games on cognition, research had explored whether video games with high spatial demands might improve performance on tasks requiring spatial imagery or spatial transformations. Part of that interest has derived from gender differences in spatial cognition and the question of whether training might minimize those differences, thereby improving female participation in domains for which such skills are important (e.g., STEM fields). Somewhat ironically, the primary game used to train spatial cognition, Tetris, is the same game often used as a control for action-game studies. In Tetris, players must rotate blocks of different shapes to form unbroken lines at the bottom of the screen.

In one of the first studies to explore possible benefits of game play on spatial cognition, participants were randomly assigned to a group that played 6 hours of Tetris or to a no-contact control group (Okagaki & Frensch, 1994). The Tetris group showed minimal benefits on the four measures of spatial ability or perceptual speed. Follow-up analyses showed a Tetris advantage specific to men on two of the measures (cube-comparison test, form-boards test), suggesting some possible selective benefits. A second experiment used outcome measures in which stimuli and presentations were more similar to those of Tetris and observed transfer to a mental-rotation task and a visualization task with no differences between men and women. Given the lack of consistency of the sex difference, the use of a passive control group, the use of multiple measures with inconsistent effects, and the presence of transfer only when the outcome measures were most similar to the training task, this study provides only limited evidence for improvements in spatial ability as a result of playing Tetris. A subsequent study with children (De Lisi & Wolford, 2002) did find greater improvements in mental-rotation accuracy after playing Tetris than after playing Where in the U.S.A. is Carmen Sandiego? (a game lacking high spatial demands).

Two more recent studies (Cherney, 2008; Cherney, Bersted, & Smetter, 2014) explored spatial training using video games. Cherney (2008) randomly assigned 61 college-aged participants to a group that played 4 hours of either Tetris or Antz (a 3-D racing game) or to a control group that completed paper-and-pencil puzzle games. The paper measured improvements on two mental-rotation measures. It reported greater mental-rotation improvements for women than men (regardless of condition), but it provided no statistical support for a difference between the two game conditions or for an interaction between participant sex and game type. When the analyses was further restricted to those participants with high mathematical ability (based on a median split), a regression analysis revealed a significant group difference, but this exploratory analysis should be interpreted with caution. Cherney et al. (2014) randomly assigned participants to a group that played 1 hour of a racing game (Crazy Taxi), a group that played 1 hour of a balance game on the Nintendo Wii (Segway Circuit), or a control group that solved paper-and-pencil puzzles. Compared to the control group, the combined game groups showed greater mental-rotation improvements for women but not for men. Given the minimal training and the fact that post-training testing occurred immediately after game play, the observed effects could reasonably reflect factors other than improved mental rotation (e.g., differential arousal or positive mood).

Some studies found no benefit of Tetris training, noting instead that the skills learned during Tetris play were specific to the context of the game (Sims & Mayer, 2002). Consistent with that pattern, Boot et al. (2008) found benefits for Tetris on mental rotation using blocks that were highly similar to those in Tetris. In a study comparing training with 2-D and 3-D versions of Tetris, only 3-D Tetris training improved 3-D mental-rotation performance at transfer, but both the 2-D and the 3-D versions transferred to 2-D mental-rotation skill (Moreau, 2013). However, this study lacked a non-Tetris control condition to rule out possible test-retest improvements on the 2-D outcome measure.

Compared to the evidence for effects of action video games on cognition, the evidence that Tetris and other games with heavy spatial demands can improve spatial abilities is more mixed, with important qualifications to reported effects and with a tendency for these studies to compare game-training groups to passive control groups.

Training benefits from strategy video games

Two studies explored whether practicing the real-time strategy game Rise of Nations might compensate for cognitive decline in aging (Basak, Boot, Voss, & Kramer, 2008; Strenziok et al., 2014). This relatively slow-paced game involves building a civilization by collecting and managing resources, although it also has some action-like components during battles. Basak et al. (2008) randomly assigned participants to play 23.5 hours of Rise of Nations over the course of 8 weeks and compared their performance to that of a no-contact control group. Both groups were assessed before and after training, as well as at the midpoint of training, on 10 cognitive outcome measures assessing executive control and visuospatial processing. The trained group improved more than the control group on four out of five measures of executive control and one out of five measures of visuospatial processing. Rise of Nations score improvements also were correlated with cognitive improvements on the outcome measures, although interpreting such correlations between training and transfer improvements is problematic (Tidwell et al., 2014). In a similar intervention, Strenziok et al. (2014) found limited transfer from Rise of Nations to memory, reasoning, and everyday-ability measures. This study found no transfer to untrained cognitive functions as a result of Rise of Nations training, but instead found some transfer from auditory training with Posit Science Brain Fitness and from training with a game developed by cognitive psychologists called Space Fortress, which is used as a tool to understand the effectiveness of different training strategies (Donchin, Fabiani, & Sanders, 1989). One limitation of this study is that practice on all of these games might have resulted in some improvements, masking potential benefits from Rise of Nations. These studies provide limited evidence that strategy-game training produces generalizable benefits.

Training benefits from casual video games

Commercially available casual video games are among the newest to be studied as possible interventions to improve cognition. Unlike action games, casual games typically are easy to learn, require minimal time commitment for each gaming session, and often are designed for phones or tablets (e.g., Angry Birds). Given the wide variety of such games, successful performance can tap many cognitive abilities (Baniqued et al., 2013). Does extensive experience playing such games sharpen those abilities such that the improvements generalize to other tasks? The few studies of this approach to training have produced mixed results.

Young adults (N = 209) assigned to play 15 hours of games that either did or did not tap reasoning and memory abilities showed no differential improvements on related measures of transfer (Baniqued et al., 2013). In another, smaller study (N = 55), Oei and Patterson (2014) reported that participants who played Cut the Rope for 20 hours showed bigger improvements on three measures of executive control than did participants who played a more fast-paced casual game (Fruit Ninja), a real-time strategy game (Starfront Collision), or an action game (Modern Combat). However, the statistics reported in the paper did not support that claim of differential improvements: The Time × Condition interaction was not significant for any of the outcome measures (all p values between .25 and .86). Another study by the same research group (Oei & Patterson, 2013) compared groups trained for 20 hours on one of five games (ns of 14–16 per group). Those trained with an action video game improved more on an attentional blink task. However, the findings reported in the paper make it difficult to determine whether or not any group showed differential improvements. The only analysis reporting a significant Time × Condition interaction included all five groups, and the remaining analyses reported improvements only within groups. In order to claim differential improvements, the improvement in a trained group must be compared directly to that in a baseline group. Otherwise, it is impossible to determine whether the active ingredient in that game was responsible for the improvement.

In summary, casual video games are diverse, potentially covering a wide range of cognitive abilities. Given the limited research to date, there is not enough evidence to conclude with confidence whether or not casual gaming (or playing specific casual games) improves cognition.

Training benefits from Nintendo Brain Age and Big Brain Academy

Nintendo's Brain Age, first released in 2005, was one of the first successful mass-marketed brain-training products. The program, also known as Dr. Kawashima's Brain Training, was based largely on a 2003 book of puzzles and math exercises authored by neuroscientist Ryuta Kawashima (who appears as a character in the game). Brain Age and its variants feature mini-games that require players to complete math problems quickly, read aloud, count syllables in phrases, count objects on the screen, or perform other spatial, verbal, and arithmetic tasks. A few published studies have examined the effects of playing Brain Age on cognitive outcome measures.

The two studies reporting the largest benefits were conducted by Kawashima and his colleagues (Nouchi et al., 2012; Nouchi et al., 2013). Nouchi et al. (2012) compared the effects of a total of approximately 5 hours of training with either Brain Age or Tetris in a sample of older adults (n = 14 in each group). The Brain Age group showed bigger improvements on two measures of processing speed and two measures of executive function, but no improvements on measures of attention, memory, or global cognitive status. The analyses did not correct for multiple comparisons, and the authors did not report results without the many included covariates. The authors acknowledged that, given the limitations of the study, "long-term effects and relevance for everyday functioning remain uncertain as yet." Benefits of Brain Age training were also observed for younger adults across a wider range of measures in the other study by the same group, which used a similar design (Nouchi et al., 2013).

Other studies of near transfer from Brain Age have been mixed. For example, Brain Age training produced greater improvements in a digit-span task when compared to a passive control, but it produced no differential improvement in arithmetic performance, even though Brain Age specifically focuses on calculations (McDougall & House, 2012). In a large-scale intervention involving schoolchildren (N = 634; D. J. Miller & Robertson, 2011), students who received 15 hours of Brain Age training improved more than those in a passive control group in arithmetic accuracy, although the effect was small (η_p^2 = .01) and would not have survived correction for multiple comparisons. A larger difference was observed for arithmetic speed (η_p^2 = .12). Thus, Brain Age might produce some near transfer to similar outcome measures, but the pattern of transfer is inconsistent, and training does not reliably generalize to tasks with demands similar to those of the training tasks.

Big Brain Academy is a game similar to Brain Age and includes exercises involving reasoning, memory, math, and other skills. In a study testing the efficacy of Big Brain Academy training (Ackerman, Kanfer, & Calderwood, 2010), 78 older adults completed two activities: 20 hours of Big Brain Academy training and 20 hours of a reading-based control activity (in counterbalanced order). Unlike almost all other brain-training studies, which have compared different participants in the training and control conditions, in this study, participants served as their own control. Reasoning and processing-speed outcome measures were assessed at the beginning of the study, between the two 20-hour treatments, and at the end of the study. Although participants gained domain-specific knowledge from the reading materials and improved their game performance within Big Brain Academy, neither training condition produced differential transfer on the primary outcome measures.

There are reasons to be cautious about interpreting the evidence from Brain Age interventions (in addition to the inconsistent patterns of results for similar training games). The game's continuous feedback in the form of an improving "brain age" could establish and reinforce differential expectations for improvement. Consistent with this concern, participants in a Brain Age intervention had significantly stronger beliefs that the game had the potential to improve their everyday abilities (Boot, Champion, et al., 2013). Currently, the evidence that Brain Age and its variants can meaningfully improve cognition is relatively weak, with few studies examining the effect of training on real-world, everyday outcome measures, and with studies featuring comparisons to control groups that did not control for potential placebo effects resulting from continuous feedback about how the game was improving participants' cognitive function.

Summary

Overall, the evidence that video-game training can improve cognition is mixed. Studies of the benefits of action-game training are among the best designed in the brain-training literature, typically contrasting performance in an action-game training group to that in an active control group with fairly well-matched task demands. Unfortunately, underreporting of outcome measures and possible non-independence of published results compromise the strength of the observed benefits. As is the case for almost all brain-training studies, these action-game studies typically have not examined whether the control group and the intervention group had comparable expectations for improvement for each outcome measure. Also, tests of those expectations suggest that the observed differences in some outcome measures between the action-game condition and the control-game condition are consistent with the pattern of expectations that people have for such improvements, raising a concern about differential expectations/placebo effects (Boot, Simons, et al., 2013). Furthermore, several failures to observe benefits, coupled with possible publication bias in the literature, raise concerns that the reported effects might be weaker than initially thought or that unknown moderators might determine whether training will be effective for an individual. No action-game-training study has tested for transfer of training to meaningful everyday tasks such as driving.

Evidence linking training on Tetris, strategy games, casual games, and Brain Age to cognitive benefits is even more tenuous. These studies have often relied on passive control groups, and they have produced inconsistent and relatively weak support for game-training benefits.

Discussion

The evidence cited by Cognitive Training Data and by brain-training companies, together with evidence from working-memory training and video-game-training studies, provides little compelling evidence for broad transfer of training. Consistent with earlier evidence dating back to Thorndike's (1906) studies in the early 20th century, training tends to produce relatively narrow transfer but not broad transfer.

Across the literature, many studies have shown benefits of training on closely related tasks. For example, speed-of-processing training using the UFOV often produces benefits for closely related laboratory measures of processing speed, and working-memory training can produce benefits for closely related memory tasks. Few studies provide evidence for transfer from one cognitive domain to another. The largest randomized controlled trial, ACTIVE, provides evidence consistent with this pattern, largely showing improvements within a trained domain but not across domains.

Few studies in this literature have objectively measured improvements in real-world performance, instead using other laboratory tasks or neuropsychological batteries as outcome measures. When studies have measured real-world outcomes, the measures have tended to be self-report and subject to demand characteristics and expectation effects. None of the published studies in this literature provide compelling evidence consistent with broad-based, real-world cognitive benefits from brain-training interventions. Perhaps the closest evidence for such benefits comes from the ACTIVE trial, which showed some evidence for reduced driving crash risk in the speed-training cohort relative to the no-contact control group (Ball et al., 2010). Still, other studies directly exploring the immediate effects of speed training on driving performance do not provide compelling evidence for benefits (Roenker et al., 2003; see also Simons, 2010).

Few of the studies we reviewed conformed to best practices for the design and reporting of intervention research. Most used inadequate control groups; almost none tested for differential expectations for improvement in the training and control groups (differential placebo effects); many neglected to report all outcome measures; few corrected statistical significance tests for multiple comparisons; some did not report the critical Time × Group interaction; and some reported outcomes from the same intervention across multiple papers without acknowledging the lack of independence. Many of the studies tested small samples, limiting their ability to detect small effects and increasing the chances that any observed effects would be spurious (Button et al., 2013). With only a few exceptions, those studies that are registered at ClinicalTrials.Gov either were registered after data collection or did not include adequate details about the method and measures. Collectively, these suboptimal practices undermine the strength of the evidence for cognitive benefits of brain training, often rendering the results reported in individual papers or collections of papers uninterpretable.

In the early stages of a field or when exploring a new form of training, less well-controlled intervention methods might be justified. Before conducting an expensive, large-scale, randomized controlled trial for a new intervention, it makes sense to explore whether performance on the intervention tasks is correlated with the desired outcome measures or to examine skilled and unskilled performers. Even conducting a pilot intervention without a baseline (or with a historical baseline) might be justified before conducting a larger, controlled trial—if participants show no improvement from the intervention, then they are unlikely to show differential improvements relative to an appropriate control group. Although potentially valuable, such preliminary evidence should not be touted as evidence for the effectiveness of an intervention—something commonly done by brain-training companies. Many such benefits from preliminary studies may not withstand the addition of an appropriate control group.

Marketing claims and the evidence from brain-training studies

The limited evidence in the literature for transfer from brain-training interventions to real-world outcomes stands in stark contrast to the marketing claims of many brain-training companies. Most companies promote their products with claims, explicit or implied, of broad transfer to everyday activities, yet none provide compelling evidence for such transfer. Many companies quote testimonials from customers touting their product's benefits for everything from workplace success to bowling. Yet almost none of the cited studies reported tests of the benefits of training on the marketed products for objectively measured real-world performance.

A few companies focus explicitly on enhancing performance among those with cognitive impairments or delays. For example, Cogmed's training focuses mostly on children with learning disabilities, and its marketing targets that segment of the population. Yet many brain-training companies cite studies testing cognitively impaired patient populations but market their products more broadly. For example, many studies cited by Cognitive Training Data focused on cognitive remediation for people with schizophrenia. For the most part, though, companies market their products to the broader public and not to people with schizophrenia. Studies of cognitive remediation or those testing patient populations are valuable, but they do not justify claims of benefits for other populations. Even when those studies provide compelling evidence for cognitive gains from training, they provide no evidence that the same training would benefit people experiencing the more typical effects of cognitive aging or those without obvious cognitive impairments.

Even if we had found compelling evidence that brain-training products benefit real-world cognition, such benefits must be considered in light of alternative uses of the time and resources they require. That is, before starting a brain-training protocol, people should consider the comparative effectiveness of that protocol. Few studies in this literature compared the relative effectiveness of different brain-training products in well-designed, randomized controlled trials.

Comparative effectiveness and appropriate inferences

Evaluating the practical usefulness of a brain-training intervention involves several questions: Does the intervention provide compelling evidence for the causal potency of the intervention? Do those benefits apply to real-world performance? Are the benefits worth the time, effort, and costs necessary to achieve them? Up to this point, we have not considered the final question. To answer it, we must consider the comparative effectiveness of different interventions as well as the value of any training benefits.

In the health sciences, determining comparative effectiveness involves weighing both the potential benefits and harms of a treatment relative to other types of treatment (or non-treatment). In the brain-training literature, direct harms are less of a concern. None of the registered clinical trials have reported any adverse events from their training, which is unsurprising, given that the interventions largely involved performing tasks on computers. Attrition rates from large studies might suggest a possible harm, but it is more likely that most people leave interventions because their personal cost-benefit assessments weigh against continuing in the study.

In the absence of direct harm or risks, determining comparative effectiveness involves first asking how well an intervention works relative to other interventions. For example, if one brain-training regimen yields a 10-point improvement on an outcome measure and another intervention yields a 12-point improvement, we can conclude that the second intervention was 20% more effective. Such comparisons are just the first step in evaluating which interventions have the most benefit. All else being equal, choosing the intervention with bigger benefits makes sense: The opportunity you chose is better than the alternative. However, all else rarely is equal. Choosing to undergo an intervention has an "opportunity cost"—your time and money could be spent doing something else, including something other than a brain-training intervention.

In order to measure the utility of a brain-training intervention, you must consider not only the relative benefits of different interventions but also their opportunity costs. If the hypothetical brain-training intervention that yielded 20% better performance also took twice as long to complete, that additional benefit might not be worthwhile. If a brain-training program produces little benefit, the opportunity costs can be large—your time and money could be used for other things. Such costs might differentially affect older adults on fixed incomes who have a shorter time horizon. Of course, the intervention could have benefits in terms of entertainment and enjoyment even if it provides no cognitive benefits, so all types of benefits should be considered when evaluating the opportunity costs.

When evaluating opportunity costs, you should also consider the cost effectiveness of an intervention. For every dollar (and/or unit of time) invested, how much benefit do you get? If an intervention costs $100, takes 20 hours to complete, and increases performance by 10 points on an outcome measure, it costs $10 and 2 hours for each point. If you calculate the value of your time to be $20 per hour, then the total cost for each point on the outcome measure is $50, and the total cost for the benefit accrued is $500. The question you must ask yourself is whether that improvement justifies the cost or whether that time and money could be better spent. Again, when an intervention has no benefits of any sort, spending any time or money on the task is wasting opportunities.

For most brain-training interventions, the benefits are not described in terms that an individual can translate directly into personal benefits. Instead, they might be stated in terms of percentage improvements on a composite measure of laboratory tasks. In order to help consumers evaluate whether or not to complete a brain-training program, these findings should be reframed in practical terms that are conceptually equivalent to dollars-per-point calculations.

For a practical example, we will revisit the evidence from ACTIVE that speed-of-processing training decreased at-fault crash risks for older adults. The question potential brain-training users would really like to have answered is how many fewer crashes they should expect to have—compared to if they chose not to undergo training. Or, how much less likely would you be to get in a crash? Given that people drive different numbers of miles, you might want to know how much your per-mile chance of a crash would decrease as a result of training. You also might want to know whether these additional crashes would be severe enough to cause injuries or death.

The ACTIVE trial estimated the number of *at-fault* crashes per year to be 0.035, 0.030, 0.023, and 0.019 per participant in the control, memory, reasoning, and speed training conditions, respectively (Ball et al., 2010). Thus, the speed-training group had a roughly 50% reduction in crash rates relative to the control group, a statistic touted by brain-training companies. Yet what people really want to know is whether they are less likely to experience a crash as a result of having undergone speed training. These numbers show a difference of 0.016 crashes per year, on average (0.035 – 0.019). So, for the typical individual, the benefit works out to roughly one fewer at-fault crash every 62.5 years. A 50% reduction sounds big. One crash every 62 years might not.

Consumers should also consider whether the reported outcome is the one they care about. In terms of risk of harm, it probably does not matter to you whether or not you caused the crash. From the same data, the overall crash rates, including both at-fault and not-at-fault accidents, were 0.043, 0.033, 0.031, and 0.038 for the control, memory, reasoning, and speed-training conditions. Ten hours of speed training reduced the expected number of crashes in the trained group relative to the control group by .005 crashes per year. In other words, a typical person undergoing speed training would expect one fewer crash every 200 years. The benefit was actually somewhat bigger for the other training conditions, although still on the order of a one-crash difference for every 100 years.

Moreover, if the overall crash rates are similar for the speed-training group and the control group, the difference for at-fault crashes must be offset by a difference in not-at-fault crashes in the opposite direction. In fact, the same data show that the speed-training group had 0.018 not-at-fault crashes whereas the control group had 0.008. In other words, the group that underwent speed training was more than twice as likely to be in a crash caused by someone else, a bigger difference than that observed for at-fault crashes. You will not see a brain-training company tout the finding that speed training more than doubles your risk of being in a crash caused by someone else. Yet that conclusion follows just as well from the data as the claim of a 50% reduction in at-fault crashes. The better conclusion might be that the difference between at-fault and not-at-fault accidents is likely just noise due to the small number of crashes in this sample—it is not particularly informative. From a consumer's perspective, the most relevant statistic is the probability of avoiding all crashes, and that does not vary meaningfully for a typical individual as a function of training.

From a practical perspective, speed training has no significant effect on your risk of a crash up to 6 years after training. Drivers might better reduce their risk by minimizing unnecessary driving and by reducing the amount of driving they do at night or in bad weather (Rothe, 1990).

These analyses are based on a comparison to doing nothing at all, but a better analysis of the efficacy of speed training for driving crash risk reduction would also consider the opportunity costs: Might some other form of training yield bigger benefits if your primary goal is to reduce crashes? For example, elderly adults often struggle when making left turns across traffic, a situation that arises in many crashes involving elderly drivers (Stutts, Martell, & Staplin, 2009). Perhaps a 10-hour training regimen focused on practicing left turns would lead to bigger reductions in crash risks. Or perhaps educational training on strategies to avoid situations that are more likely to pose challenges for older drivers would produce greater benefits (e.g., making three right turns instead of a left turn).

Comparing the effectiveness of interventions is complicated by possible differences in the relationship between the training dose and the measured effectiveness. The power law of practice (e.g., Heathcote, Brown, & Mewhort, 2000; Newell & Rosenbloom, 1981) suggests that training gains generally will be rapid at first and then diminish rapidly. Consequently, increasing the amount of training might produce diminishing returns. Almost no studies have examined the shape of these dose-benefit functions or compared them across interventions, and most have chosen an arbitrary total amount of training. Consequently, consumers cannot directly compare the effectiveness of the same amounts of training on different interventions.

Consumers also likely want to know whether any benefits of training will last, and the persistence of the effects should factor into their decision about opportunity costs. A benefit lasting for years would be more valuable than one lasting for weeks. With the exception of a few of the largest interventions (e.g., ACTIVE), few brain-training studies have examined the persistence of training benefits, and those that have typically have not examined whether the dose of training affects its longevity. Across all of the studies we have reviewed, there is little evidence for long-lasting benefits on tasks other than the trained ones.

Finally, consumers want to know if an intervention will work for them, personally. Interventions might not affect each individual in the same way. Just as with physical exercise, some people might show little benefit and others might experience robust benefits of any intervention (Bouchard et al., 2011). Tailoring training to account for individual differences would be ideal, but the field lacks the large-scale studies necessary to determine whether such individual-difference factors mediate or moderate the effectiveness of training.

In order to provide effective guidance to the public, we need assessments of the effectiveness of the training itself (like our analysis in this article), but we also need studies assessing the comparative effectiveness of interventions that do work. Moreover, we need to consider the opportunity costs and generalizability of those interventions. At present, none of those further analyses are possible given the published literature.

Recommendations

In this section, we recommend a set of aspirational standards for scientists, journalists, policymakers, and consumers interested in brain training. These recommendations are intended to improve the quality of studies testing promising new approaches to cognitive training and to help consumers understand the likely costs and benefits of such interventions.

Recommendations for scientists

This section highlights ways to improve the quality of future studies so that they will yield more definitive conclusions about the effectiveness of brain-training interventions. Those recommendations include guidelines for designing research and reporting results. Reviewers and editors have a responsibility to verify and enforce best practices so that publications do not mislead the scientific community or the public. To that end, we provide a list of questions that reviewers and editors should ask before accepting a paper based on an intervention study (Table 5).

Recommendations for theoretical development. Occasionally, a new intervention proves successful before researchers fully understand why it works; although transfer of training has been studied for more than 100 years, the mechanisms that can and should produce transfer are still not well understood. The brain-training literature is grounded mostly in implicit theories of how cognitive capacities support everyday functioning. For example, Posit Science assumes that representational fidelity and speed of processing are tightly linked to the performance of many everyday tasks, and Cogmed assumes the same about working memory.

For these assumptions to become grounded, researchers must explain, in a mechanistic way, why their particular training approach should enhance cognition broadly when most cognitive training yields only narrow benefits. Yet the field lacks a full understanding of the relationship between the training tasks and the targeted outcome measures or even among the tasks themselves. For example, *n*-back and complex span tasks (e.g., operation span) that are commonly used to measure working memory are only weakly correlated and make independent contributions to measures of fluid intelligence (Kane, Conway, Miura, & Colflesh, 2007). Without a full understanding of the mechanisms underlying working memory or how these mechanisms contribute to other aspects of cognition, the choice of a working-memory task as the basis for an intervention involves implicit assumptions about which ingredients are key. Moreover, just because two (or more) cognitive tasks are correlated (most cognitive tasks are; Deary, 2012) does not mean that training on one will enhance performance on another (e.g., Ball et al., 2002; Willis et al., 2006). More broadly, the field must come to a more complete understanding of the relationships and dependencies among both laboratory tasks and real-world performance in order to optimize training regimens.

Without a clear understanding of what it means to “help you academically, socially, and professionally,” the goals of an intervention are underspecified. An analogous clinical trial in medicine would have a vague desired outcome like “improved health.” The challenge is to identify well-defined outcome measures that also demonstrate causal links to everyday cognitive performance. The larger challenge involves identifying and validating objective measures of those everyday cognitive activities targeted by an intervention. For example, how do we measure improvements in academic, social, or professional performance? Each of these complex domains likely varies along many measurable dimensions, each of which could be targeted by an intervention. More broadly, cognitive-intervention research needs more complete translational theories that meaningfully connect lab-based measures to objective measures of everyday performance.

When a measure—say, IQ—correlates with real-world outcomes such as work performance, some might be tempted to conclude that improvements in IQ necessarily will translate to improvements in work performance. So, for example, if working-memory training led to improved

Table 5. Questions Reviewers and Editors Should Ask Authors, and Appropriate Responses When They Answer "No"

Question	If "no"
Was the study preregistered, and did that preregistration include full details about the design and analysis plan?	Were the results appropriately described as exploratory?
Did the paper include a statement like, "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures"?	Ask the authors to update the manuscript until they can answer "yes" or satisfactorily explain why they cannot. We recommend adding the following standard disclosure request to any review that lacks such a statement: "I request that the authors add a statement to the paper confirming whether, for all experiments, they have reported all measures, conditions, and data exclusions and how they determined their sample sizes." The authors should, of course, add any additional text to ensure the statement is accurate. This is the standard reviewer disclosure request endorsed by the Center for Open Science (see http://osf.io/hadz3).
Is all data collection for this phase of the study completed?	Wait to publish the paper.
Does the paper make clear which analyses are confirmatory and which are exploratory?	Insist that only preregistered hypothesis tests be treated as confirmatory and that others be identified as exploratory.
Does the paper include a CONSORT diagram with the numbers of participants recruited and tested in each condition?	Require one.
Does the paper include a table showing all outcome measures with descriptive statistics for each condition both before and after training?	Require one.
Does the paper fully describe all training tasks and their relationship to the outcome measures?	Require a complete description or a reference to another source with a complete description.
Does the paper state clearly that participants were (or were not) randomly assigned to conditions?	Require a clear statement of how participants were recruited and assigned to conditions.
Have any other papers been published based on data from these same participants and this intervention?	Require full documentation and justification of the overlap. Require correction for multiple testing within and across papers if needed.
Did the paper include an active control group that was suitably matched to the intervention group?	Require the authors to temper any claims about the effectiveness of the intervention and to acknowledge all of the factors other than the critical ingredient of the intervention that might contribute to differential improvements.
Did random assignment lead to equivalent baseline performance across groups?	Differential improvements on a measure that started with different baseline performance are difficult to interpret. Any claims about them should be qualified by noting the possibility of a failure of random assignment to equate the groups on that measure.
Does the paper report a focused test of the differential improvement across conditions for each outcome measure?	Require a direct assessment of the improvement in the intervention group relative to the control group.
Did the study test whether expectations for improvement were comparable for each outcome measure across groups?	Either require such a check or require an acknowledgment that differential placebo effects could explain differences between conditions.
Have the authors made the data from their intervention publicly accessible?	Ask them to do so or to explain why they cannot.
Do the authors list any connections (financial or otherwise) with brain-training companies?	Ask them if they have any connections or conflicts of any sort and require that those be acknowledged in the paper.
Do the authors fully acknowledge the limits of the inferences that follow from their intervention design?	Ensure that the authors acknowledge all factors other than the intervention that could contribute to differential improvements between the intervention and control condition.

IQ scores, researchers might be tempted to conclude that the training would benefit all of the activities associated with higher IQ. However, the aspects of IQ that drive its predictive validity might not be the same as those influenced by training. Moreover, the elements of a task that benefit from practice might not be those that are correlated with real-world outcomes; practicing an IQ test might improve your performance on IQ tests while simultaneously reducing the validity of IQ tests in predicting real-world outcomes. In order to claim a benefit of training for real-world performance, it is essential to measure that effect directly rather than basing it on improvements in correlated measures. By analogy, if you want to test whether a drug effectively treats diabetes, you must test its effect on diabetes, not on other factors that correlate with diabetes. In order to make claims about real-world benefits, interventions should assess real-world performance. And, when possible, they should test how long the observed benefits persist after training.

In addition to a more fully developed understanding of the relationship among abilities, cognitive-training interventions would benefit from a richer understanding of differences between the populations under study. Many brain-training products are marketed to healthy adults who suffer no measurable deficits but want to maintain or enhance their cognitive performance. Yet many intervention studies cited by brain-training companies as support for their claims have tested populations with notable cognitive deficits due to factors such as age-related pathology, schizophrenia, depression, or brain injury. These problems have distinct etiologies, and companies rarely explain why interventions to remediate cognitive impairments in one population necessarily will help people from other populations.

Intervention research need not wait for a fully elaborated theory, of course. Effective interventions regularly emerge in the absence of an elaborated, complete theory, and such interventions can inform subsequent theorizing. The mechanisms for many medical interventions, especially in domains such as mental health, were not understood when first introduced (e.g., SSRIs for the treatment of depression, Stahl, 1998; lithium for the treatment of bipolar disorder, Lenox, McNamara, Papke, & Manji, 1999). In most such cases, the symptoms are well defined, even if the mechanisms remain opaque. Yet treating the symptoms without understanding the mechanisms can lead to unforeseen consequences; thalidomide was an effective treatment for nausea symptoms during pregnancy but had horrific and unforeseen consequences for fetal development.

Cognitive interventions, even in the absence of a mechanistic understanding, lack the risks of medical ones. Laboratory research on learning, transfer of training, and the links between cognitive tasks and real-world performance underpins the development of such theories. And tests of the efficacy of a brain-training product or intervention can precede theory. However, conclusions about *why* such interventions are effective or *which* mechanisms have improved depend on a more complete theory.

Recommendations for study design and documentation. Our review has identified a number of design and analysis shortcomings common to brain-training interventions. Here, we provide recommendations, based on current best practices, for improving the quality of future brain-training interventions. Such standards evolve over time, but even if better approaches emerge, the guidelines we provide would eliminate some substandard practices, yielding better and often more powerful tests. Several of these recommendations, such as using larger sample sizes and reasonable amounts of training, should increase the chances of finding benefits if an intervention really does yield transfer of training.

Some recommendations, including preregistration, should be followed by all intervention research. Other recommendations may not be possible during the early, discovery stage for new interventions or may be infeasible for researchers with limited resources. Studies that cannot implement these best practices, however, must acknowledge those limitations and restrict their claims and conclusions accordingly. Direct replication of such limited findings, using a preregistered design and analysis plan, can help ensure the robustness of any observed improvements.

Preregistration. All future intervention research should be registered prior to the start of data collection. This preregistration should specify all of the following: the target sample size, the characteristics of the tested population, the nature of each intervention group, the technique for randomly assigning participants to conditions, the rules for excluding data, a list and description of all outcome measures, a full description of the a priori hypotheses, and a complete analysis plan for testing those hypotheses. The registration should include all testing materials and experimental scripts used to conduct the study and all analysis scripts that will be used to analyze the data.

Preregistration is a feature of good design because it makes clear how the study will be conducted and what can be learned from it. Many websites make preregistration straightforward (e.g., ClinicalTrials.Gov, Open Science Framework). Preregistration can also take the form of a "method" paper published in advance of conducting the study, although website-based preregistration is more flexible and easier to access and update. Preregistration does not preclude exploration—studies can include additional measures that are not essential for a priori hypothesis tests, and researchers should fully explore their data. Preregistration makes transparent and unambiguous which analyses constitute confirmatory hypothesis tests and which are part of a broader discovery process.

Publications should cite a preregistered design and analysis plan prominently and should indicate and justify any deviations from that plan. Any analyses or measures that were not part of a preregistered, confirmatory hypothesis test should be explicitly labeled as exploratory, and the paper should make clear that readers should take those findings as tentative until they have been replicated by an independent study designed to test exactly that question. Such exploratory analyses should not be used as the basis for recommending actions on the parts of readers, therapists, or users of any brain-training product.

Randomization and sampling. Beyond the pilot stage, studies should test a large number of participants and should randomly assign them to conditions. They should describe in detail the procedures used to randomly assign participants to conditions. Effective randomization requires large samples—with small samples, randomization can introduce sample differences between conditions that could undermine the ability to observe training effects or spuriously increase them. When relatively small samples must be used, matching participants on known confounding factors (e.g., age, intelligence) and then randomly assigning the members of each pair to conditions is an alternative approach (C. S. Green, Strobach, & Schubert, 2014). Better, though, would be to test a large number of participants by combining efforts across laboratories. Studies with small numbers of participants should be viewed with greater skepticism and should be treated as tentative.

Sample sizes should be large enough to reliably observe effects of a reasonable size, and a priori power analyses should be conducted to document that the study design is sufficiently powered. As a benchmark, imagine a study with one treatment group and one control group, each tested before and after an intervention. If the study tested just one reliable outcome measure (.75 correlation between pretest and posttest), it would require more than 100 participants to have 80% power to detect a small ($f = 0.10$) Time × Condition interaction. For a less reliable outcome measure (e.g., a .50 correlation between pretest and posttest), the study would need 200 participants. If the study had five outcome measures with that reliability, after correcting for multiple comparisons ($\alpha = .01$), it would need nearly 300 participants. Studies should justify why a chosen sample size is sufficient to detect the differential improvement between conditions after correcting for multiple outcome measures. If doing so requires assumptions about the reliability of the measures and the size of the expected effects, those assumptions should be justified.

Dosing. Researchers conducting intervention studies should justify their chosen amount of training. Larger amounts of training might be more likely to reveal benefits, but they also impose a larger cost on the trainee. If improvements require a large amount of training, they might be less effective when introduced to the public (as a result of reduced adherence). For training benefits that do require a large dose, researchers should discuss the comparative effectiveness and costs of that intervention.

Blinding. In most brain-training interventions, participants cannot be blind to the nature of the intervention; they know which tasks they are doing. Consequently, they may form expectations about the links between their intervention tasks and the outcome measures. When those expectations align with the pattern of results, any differential improvements might reflect a placebo effect rather than the effectiveness of the intervention itself. To address that issue, study designs must take care to ensure that participants and researchers are as blinded as possible to any factors that might lead to differential outcomes. Participants should be blind to the nature of the other conditions and, if possible (pragmatically and ethically), to the existence of other conditions. Experimenters involved in data collection for the outcome measures should be blind to the condition assignment for each participant.

Control conditions. All intervention studies should include a baseline control condition—ideally an active one. Interventions relying solely on a passive baseline condition cannot yield clear answers about the effectiveness of an intervention because too many factors differ between any intervention and a passive control condition. Interventions without matched, active control groups can only provide evidence that some difference between the intervention and control conditions contributed to the differences on outcome measures. Including a no-contact control group as well can provide a baseline for the effects of retesting, but the active comparison group is more critical for inferences about the causal potency of an intervention. The active control condition should be as similar as possible to the training condition, excepting only the ingredient thought to be central to the training benefits. For example, a video-game intervention should use an active control condition that also involves gaming, just not those aspects of the game that are thought to produce the benefits.

In addition to matching the demands of the training condition to those of an active control condition, intervention studies should evaluate whether expectations for improvement are aligned with any differential improvements across groups. If they are, then the different outcomes for the intervention and control groups might be explained by a differential placebo effect, and investigators should acknowledge that possibility. Critically, these expectation checks are necessary for each outcome measure. Overall measures of motivation are helpful, but they do not account for differential placebo effects on each outcome measure. Measuring expectations in the trained participants can provide some information, but doing so may sensitize them to the manipulation and could actually induce expectation effects. Measuring expectations for the effects of a training intervention or control task among a separate group of participants is a way to address this concern (e.g., Stothart, Simons, Boot, & Kramer, 2014).

When expectations for differential improvements across conditions are distinct from the actual pattern of improvements (e.g., the group expected to improve more actually improves less), we can be confident that those differential improvements were not driven by a placebo effect. When expectations and outcomes are aligned, however, it becomes more difficult to disentangle a placebo effect from a treatment benefit. A differential placebo effect could entirely account for the improvements, or it might play no role at all. Expectancy effects might reflect the participants' ability to monitor their own learning or to judge what is likely to help, with no causal impact of those expectations on the intervention itself. Or, both the treatment and expectations could play a role, with expectations enhancing the effectiveness of the intervention in the trained group relative to the control group. In such cases, the relative contributions of expectations and treatments are inherently ambiguous, and follow-up research with additional control groups is necessary before marketing any product based on the effectiveness of the intervention itself.

Outcome measures. Studies should use outcome measures with known or measured reliability and should report that reliability. If outcome measures are unreliable, then improvements on those measures may be driven by factors unrelated to the intervention. Unreliable measures might also make it harder to find evidence for brain-training effectiveness.

Multiple measures should be used to assess each outcome construct. For example, if a study hopes to observe improvements in working memory, it should use multiple distinct measures of working memory (e.g., span measures, *n*-back measures). Given that individual measures of any construct may fail to capture all aspects of that construct (the "process impurity" problem), and given that individual measures are often somewhat unreliable, studies that measure latent constructs by using multiple independent measures can provide a more powerful test of the effect of an intervention on that construct. This approach can also serve to reduce the number of statistical hypothesis tests. The plans for constructing composite measures should be specified as part of the preregistration plan, before inspecting the data.

Moreover, papers should report all outcome measures separately, in addition to any latent-construct analyses. Although testing latent constructs can aid in developing and testing theories, they can be difficult to interpret when evaluating the practical benefits of an intervention. Some "observed" variables can safely be reported without examining latent constructs (e.g., mortality, the need for institutionalization, crash rates), especially if they can yield clear policy implications. Ideally, outcomes should include objective measures of real-world performance whenever possible.

Recommendations for analysis, reporting, and publication. Many of the reviewed studies did not fully report the necessary statistical tests or the information necessary to evaluate the size of the reported benefits. Most studies have adopted a null-hypothesis significance-testing approach. Other possibly superior approaches are possible (e.g., Bayesian estimation, multilevel modeling), but we focus on recommendations for the most common approaches.

Directly compare improvements. Studies must analyze the difference in improvement between the intervention and control groups. That analysis typically takes the form of a Time × Condition interaction term in an ANOVA. An alternative and possibly better approach might involve using analysis of covariance (ANCOVA) to compare posttest scores across conditions while controlling for pretest scores. That approach accounts for variations in pretest scores both within and across conditions and, assuming that pretest performance is comparable between the treatment and control groups, it might constitute a more statistically powerful test of differential improvements as well. A significant improvement in the treatment group but not in the control group does not provide evidence that the benefit of the treatment was greater: A difference in statistical significance is not the same as a significant difference. Similarly, a comparison of posttest scores that does not control for pretest scores does not test for differential improvement. An analysis of the effectiveness of an intervention must directly test the difference in improvement between the intervention group and the control group.

Statistical tests should be focused, meaning that they have no more than 1 degree of freedom in the numerator. The interaction term in a 2 × 2 ANOVA fits this criterion, but a 3 × 2 interaction does not. A 3 (group) × 2 (time) interaction shows only that the three groups differed somehow in the effect of time. For the same reason, a 2 (group) × 3 (time) interaction cannot provide clear evidence for differential improvements in the training and control groups.

Statistical tests of the effectiveness of an intervention should include only the training group and the matched control group, and they should not include groups from different populations. Imagine a study that randomly assigned patients with brain damage to a treatment group and a control group, but added an additional group of unimpaired adults who also underwent treatment. If all three groups were included a 3 (condition) × 2 (time) ANOVA, the Time × Condition interaction might be statistically significant even if the matched treatment and control groups improved equally, undermining any claim about the effectiveness of the intervention itself. When testing the effectiveness of an intervention, the analysis should include only those conditions that allow a direct inference about the role of the treatment in differential improvements.

Correct for multiple comparisons. When using null-hypothesis significance testing and setting a fixed cutoff for statistical significance (e.g., $\alpha = .05$), it is essential to correct for multiple comparisons in order to maintain the experiment-wide false-positive rate at that cutoff level. If a study includes 10 outcome measures and tests each for differential improvements across conditions, the alpha level should be adjusted to account for the possibility that one or more of those tests might produce a spuriously significant difference. Moreover, each test in an ANOVA constitutes an independent significance test, so analyses should be corrected for each main effect and interaction F value. In essence, the need to correct for multiple comparisons means that a study testing a large battery of outcome measures must set a more stringent criterion for statistical significance in order to maintain a false-positive rate of 5%. Note that the need for corrections applies to a priori hypothesis tests but not to exploratory analyses. An advantage of preregistration is that researchers can specify in advance which analyses test core hypotheses. For exploratory analyses, a better approach might be to provide estimates of effect size rather than conducting null-hypothesis significance tests. Any finding revealed in such exploratory studies should be subjected to a confirmatory replication with a preregistered analysis plan before it is concluded that it constitutes a benefit of an intervention.

Provide data and statistics. The individual data underlying any published analyses should be made available publicly alongside the preregistration of that study so that others can verify the accuracy of the analyses. Published reports should include a CONSORT diagram indicating the nature of the recruited sample and the numbers of participants completing training.

Papers also should include a table documenting all outcome measures in addition to any latent construct measures (see Table 6). The table should include separate values (means, standard deviations, confidence intervals, sample sizes) for both pre- and posttest scores separately for each group of participants for each measure, and separately for participants who did and did not complete training. The same table could include the measured (or known) reliability of each measure as well as the correlation between pre- and posttest scores. That correlation is necessary when computing effect sizes for a repeated-measures analysis, but it almost never is reported in cognitive-intervention studies. Finally, the table could report the relevant comparison of improvements in each group. In addition to providing the full documentation necessary for meta-analyses, this table would also allow readers to identify all of the outcome measures tested in the study.

In addition to providing statistical summaries based on standard measures and effect sizes, when discussing the potential value of training, papers should describe the personal relevance of the effects. How much better will I perform on tasks like the outcome measures? For example, how much will the intervention decrease my risk of a crash? How much less likely will I be to forget my medication?

Focus on measuring and estimating evidence. Null-hypothesis significance tests do not provide a direct measure of the relative evidence for or against the benefits of an intervention. Rather, they express only the probability of a result as big or bigger than the observed one if a particular model were true (typically, if the null were true). Evidence is relative, and measuring evidence requires a comparison of models: Which is more likely to produce the observed data? Rather than focusing on statistical significance, future research could examine how much evidence the data provide in support of a training benefit relative to the lack of a training benefit (e.g., a Bayes factor). Doing so would also place an emphasis on the size of the training benefits we might expect and whether the evidence supports such an effect relative to the absence of any effect. By specifying the alternative hypothesis and evaluating the relative evidence for it, such studies would make explicit the magnitude of any expected (and observed) benefits. In addition to making such comparisons of the evidence under different models, future research could emphasize estimation, using large enough sample sizes to provide a precise estimate of the size of the observed benefit rather than testing its statistical significance relative to the null (see Wasserstein & Lazar, 2016, for recent recommendations from the American Statistical Association on appropriate and inappropriate uses of p values).

Avoid duplicate or scattershot publication. All papers based on the same intervention must be clearly identified as coming from the same intervention, and any nonindependence among a published set of papers must be

Table 6. A Model for Reporting Intervention Results

Measure	Group	Pretest	Posttest	Difference (posttest – pretest)	Training benefit
Construct 1 (reliability)	Intervention (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	Time × Group effect size (95% CI) or regression of post-test scores on group and pretest scores
	Control (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	
Measure 1 (reliability)	Intervention (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	Time × Group effect size (95% CI) or regression of post-test scores on group and pretest scores
	Control (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	
Measure 2 (reliability)	Intervention (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	Time × Group effect size (95% CI) or regression of post-test scores on group and pretest scores
	Control (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	
Measure 3 (reliability)	Intervention (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	Time × Group effect size (95% CI) or regression of post-test scores on group and pretest scores
	Control (*n*)	Score (95% CI)	Score (95% CI)	Score (95% CI) Pretest/posttest correlation Statistical test	

noted explicitly and described fully. The results of many of the intervention studies we reviewed were distributed across multiple papers, and many papers neglected to identify all of the outcome measures tested. Such under-reporting invalidates statistical tests and makes meta-analysis impossible. Papers should make clear whether their reporting is complete, ideally using a statement like "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures" (Simmons, Nelson, & Simonsohn, 2012). When evaluating a paper that lacks such a statement or a full explanation for why such details are not reported, readers should assume that the reporting is incomplete and should discount the reported findings.

It is problematic to report behavioral outcome measures that resulted from an intervention in separate papers without fully identifying all of the collected measures in all papers. Even with full specification of all outcome measures in all papers, such scattershot publication makes it difficult for readers to evaluate the strength of the evidence resulting from an intervention; readers have no way to know which of the outcome measures yielded significant results without compiling evidence across multiple papers, something that is not possible until all of the results are reported. Moreover, such scattershot publication makes it impossible to determine how many significance tests were conducted from the same study or whether the reported statistics adequately correct significance levels for multiple tests. Consequently, the individual reports provide only ambiguous evidence because they depend on future publication of other results from the same intervention. The demands on researchers to publish results and the incentives to do so rapidly undermine the value of the evidence in each of those papers. We encourage researchers to adopt higher standards for intervention reporting and to strive for complete reporting of a class of measures (e.g., behavioral outcomes) in a single paper even if other classes of measures (e.g., neuroimaging outcomes) are reported separately.

It is sometimes acceptable to report distinct aspects of an intervention across multiple papers. For example, one paper might report all of the behavioral outcomes of an intervention whereas another might report all of the brain-imaging outcome measures. A third paper might report the results of a 2-year follow-up study with the same sample. Each paper must identify all of the other measures collected and explicitly note that the results are based on the same intervention. All should refer to the preregistered design and analysis plan and should make clear how the analyses were corrected for multiple comparisons.

One problematic practice in the brain-training literature involves publishing preliminary reports of the effectiveness of an intervention on the primary outcome measures before data collection has been completed for all participants. In some disciplines, when preliminary evidence for a new treatment might provide the only possible benefit (e.g., a new Ebola treatment) and the relative risk is low, the benefits of publishing preliminary (and possibly wrong) results outweigh the risks. Brain-training research rarely fits that description, and the risks of publishing an incorrect and preliminary result typically outweigh the benefits of rapid publication. Consumers and policymakers do not necessarily know which findings are preliminary, and media coverage rarely differentiates preliminary from completed studies. Once a preliminary finding enters the literature, it persists in public consciousness even if later contradicted. The risk of such "zombie" findings outweighs the benefits of rushing to publish before the data are in. Given biases against publishing negative results, if the full data set contradicted the preliminary one, it might never appear in the literature. In such cases, the paper based on a partial data set should be retracted because it was proven to be in error by the full data set. Papers analyzing the results from a partial sample, before the full data set has been collected, should not be published because they can provide only a misleading view of the results of the intervention.

Some published pilot studies have also been presented as evidence of the effectiveness of brain training (six of the papers cited by Cognitive Training Data were pilot studies). Pilot studies are intended to test the feasibility and reliability of measures prior to a larger confirmatory study, and they should be treated as exploratory preparations for such a study, not as direct evidence for the effectiveness of brain training. Most pilot studies in the brain-training literature lack any control group. They were intended to test whether an intervention yielded any improvements or whether the intervention method worked (e.g., testing whether training over the Internet might be helpful). But they do not provide evidence for the efficacy of brain training and should not be touted as such.

Recommendations for acknowledging conflicts of interest. Given the steady growth of the market for brain-training products and the close links between intervention research and the brain-training companies basing their marketing on the outcomes of that research, greater transparency in acknowledging conflicts is needed. Psychology as a discipline has rarely had such close connections to a large, profitable industry and has not had as much experience as fields like pharmacology in addressing the many ways that corporate goals potentially conflict with scientific ones. Although researchers no doubt view their own work as objective and untainted by corporate influences, evidence from fields like medicine raise doubts about the ability to remain neutral when financial incentives are aligned with one outcome (e.g., see Bekelman, Li, & Gross, 2003; Garg et al., 2005; Perlis et al., 2005).

Although some researchers consistently and conscientiously identify their corporate ties in their publications and presentations, not all do. Eight signatories of the Cognitive Training Data letter noted a conflict of interest, defined by the letter as "having financial interests (research funding, stock options, or stocks) in the brain-training industry." Yet many others would have conflicts if that standard were broadened to include any relationship with a company that might be perceived as biasing (e.g., consulting relationships, board memberships, advisory roles). In our review, we observed cases in which conflicts reported in one paper from an intervention went unreported in other papers based on the same intervention. Researchers should consistently disclose any relationship with a company that could be *perceived* to affect their objectivity, even if they do not believe their objectivity has been compromised.

Studies conducted by a company or its employees are often a necessary first step in testing and validating an intervention, but independent validation and replication lend credibility to claims about the effectiveness of any intervention. When the vast majority of research cited by a company has been conducted by its own employees, the public should treat any claims as tentative until they have been verified by multiple independent research groups.

Recommendations for funding agencies

Some of the common design limitations of brain-training interventions might be due to funding constraints. Unlike some other types of federal grants, intervention grants often focus on a single large study. Granting agencies regularly ask those receiving funding to cut some percentage of the cost from a grant after they decide to fund it. For multi-study grants, researchers can cut one or more studies or reduce staffing support. For intervention studies, researchers can reduce the sample size, reduce the training dose, or cut one or more conditions. Each of these changes undermines study quality. Granting agencies should recognize the need to support reproducible and reliable science, which means ensuring that funded interventions can include appropriate sample sizes, training regimens, and control conditions.

Granting agencies also have a responsibility to ensure that interventions will yield interpretable results. As a condition of funding, they should demand a complete registration of the design and analysis plan prior to data collection; ensure that funded studies follow best practices; require public posting of all data resulting from the intervention; and require publication of all results, regardless of the outcome. When possible, they should support independent replication studies to determine whether intervention results are robust and reproducible.

Given the vast public interest in and use of brain-training software (reflected in, for example, the AARP survey), we would encourage funding agencies to allocate more money to independent, well-designed tests of the effectiveness of cognitive interventions, as well as to basic behavioral research on learning, plasticity, and transfer. Agencies should consider the relative benefits of funding cognitive-training studies designed to operate via transfer from brain games to real-world tasks versus more targeted, direct practice on those real-world tasks themselves. In order to understand the practical consequences of cognitive improvements, more research is needed on the ways in which cognitive skills are applied in the contexts of work, education, and daily life. More broadly, funding agencies evaluating the rationale for a brain-training intervention study should require an assessment of the opportunity costs; assuming transfer of training will yield broad benefits in a range of domains, will the size and scope of those benefits outweigh the inefficiency of training the desired performance directly? And how do the costs of the intervention compare to those of other potentially effective interventions, such as exercise?

Recommendations for journalists and the media

Journalists interested in evaluating the implications of a published study should ask themselves many of the same questions listed in Table 5. Unlike reviewers and editors, journalists are not responsible for ensuring the quality of published research. They do, however, have a responsibility to the public to verify that the claims made based on a study match the evidence provided by it. Many published, peer-reviewed studies have substantial design and analysis flaws that limit their value, but companies still use them to market their products; peer review does not guarantee that the reported effects are real or that the claims made in an article or press release are justified. When writing about an intervention study, journalists can take several steps to ensure that they and the public are not misled about the meaning of a study.

In addition to the questions highlighted in Table 5, journalists should examine how large the measured benefits are in terms relevant for an individual user. How much will Jane Public benefit directly from completing the intervention? For example, if the authors claim that the intervention reduces crash risks, how many fewer crashes should Jane expect over a 5-year period? Journalists should ask the authors of an intervention study whether they measured any real-world improvements directly and objectively. Critically, journalists should ask authors to justify the comparative costs and benefits of engaging in their intervention relative to spending the same amount of time on other activities or interventions. Journalists should question whether the sample population tested in the study limits the generality of the benefits and whether other differences between the intervention group and the control group might explain any benefits.

Recommendations for policymakers

If companies consistently promise more than their products deliver (based on the scientific evidence), governmental agencies should consider whether existing laws support increased regulation of those marketing claims. Brain-training products target some of the most vulnerable populations: children with cognitive deficits, adults experiencing cognitive declines, psychiatric patients with cognitive impairments. Such groups merit protection if products are marketed as an effective therapy. Although the use of such products may not cause direct harm, ineffective and time-consuming training may detract from other, more effective therapeutic or educational techniques. If companies market brain training as a proven therapeutic intervention, they should be required to demonstrate that the benefits exceed those garnered by a placebo effect.

Historically, funding agencies have played a central role in evaluating the relevance of scientific evidence for real-world applications and practices. From 1977 to 2013, the National Institutes of Health sponsored a Consensus Development Program to provide unbiased, evidence-based analyses of controversial medical issues (https://consensus.nih.gov), including, for example, a 2010 report on possible approaches to reducing the risks of Alzheimer's disease and cognitive decline among older adults (National Institutes of Health, 2010). Unfortunately, the program was retired in 2013, with the rationale that other governmental agencies and private organizations had emerged to take on this advisory role.

More recently, the Federal Trade Commission (FTC; 2016a) has examined advertising and marketing claims made by brain-training companies and, in several cases, has charged companies with deceptive advertising practices (Lumos Labs being the most prominent of these). These regulatory actions help to reduce overly broad marketing claims, but the FTC's mandate in these cases rests on whether individual marketing claims are backed by evidence. Companies could run afoul of FTC regulators because of overly strong claims even if their products do produce some benefits.

The What Works Clearinghouse (WWC), established in 2002 by the U.S. Department of Education, promotes "informed education decision-making by providing educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence about 'what works' in education" ("Frequently Asked Questions," n.d.). It commissions reports from external firms that conduct reviews according to detailed guidelines in a *What Works Clearinghouse Procedures and Standards Handbook*, and its reviews include coverage of some education-related brain-training programs. In some ways, its guidelines establish rigorous standards. For example, only randomized controlled trials with groups matched at baseline and with relatively low attrition are endorsed "without reservations." Unfortunately, the WWC's standards neglect other best practices, which undermines the trustworthiness of evidence for the effectiveness of interventions. For example, its standards pay little heed to the nature of the control group; many studies endorsed without reservation have featured passive control groups (e.g., treatment as usual). It also does not consider whether the testers were blind to the treatment conditions.

As of 2008, the *Handbook* included a policy requiring the disclosure of conflicts of interest, but it oddly exempted the developer of the intervention (e.g., the WWC treats a study conducted by the company that produces the intervention as devoid of any conflict). As a result, many WWC-endorsed interventions were conducted by people or companies who potentially could have profited from those endorsements. The consequences of this exception are not trivial. For example, in an assessment of the effectiveness of software used to support physician decision-making, 74% of studies in which the developer of the software was also the investigator showed improvements, compared to just 28% of those conducted by independent investigators (Garg et al., 2005).

Within the brain-training literature, the WWC evaluated the effectiveness of Fast ForWord and concluded that it showed positive effects for alphabet learning and mixed effects for comprehension (What Works Clearinghouse, 2010). Yet six of the seven studies endorsed without reservation were reports authored by Scientific Learning Corporation, the company that produces Fast ForWord and stands to profit from its adoption. These were not independent publications in peer-reviewed scientific journals.

Studies by companies can provide valuable evidence for the effectiveness of an intervention if they are conducted with rigor, but policymakers and consumers should recognize that companies have an inherent conflict of interest—they are less likely to publish findings that undermine their claims or that might impinge on their profits. Such concerns would be somewhat mitigated if all intervention studies were preregistered and published regardless of the outcome, and we encourage publication of all such efficacy studies. But until the field meets that standard and eliminates publication biases, evidence from companies should be viewed with greater skepticism in the absence of independent direct replication. We encourage the WWC to revise its standards of evidence and to explicitly identify and factor into its evaluations any conflicts of interest that could potentially influence the evidence it reviews.

Federal agencies and independent research evaluation firms can play an important role in the development of sound public policies on health and education by helping to set the standard for rigorous study design and reporting. They also play a critical role in helping clinicians, teachers, and the public evaluate conflicting research evidence. We hope agencies like the WWC will incorporate the best practices we discuss into their evaluation procedures and standards. Otherwise, they risk misleading rather than informing the public about the quality of the evidence backing an intervention. Public health agencies and policymakers should refrain from recommending that people use brain-training products until the benefits of those products for real-world performance have been established and directly replicated in independent studies that conform to best practices.

Recommendations for consumers

We found little compelling evidence that practicing cognitive tasks in brain-training products produces lasting cognitive benefits for real-world cognition. However, some training programs might well produce benefits for the trained tasks and closely related ones. Such improvements through targeted practice are well supported by the cognitive-learning and expertise literatures (Ericsson, 2006; Newell & Rosenbloom, 1981). The appeal of brain-training products is the potential to improve cognition in a wide variety of situations by practicing one or a few cognitive tasks. But relying on transfer and broad

Problems With Intervention Studies and Their Implications

Throughout this article, we have identified many design and analysis problems common to brain-training interventions. Although all of these problems can weaken the evidence for brain-training benefits, some are more troubling than others. Below, we classify these problems into different categories based on their severity. Even if a study avoids all of these pitfalls and problems, that does not make it definitive. Individual studies can provide spurious results, even if they adopt all best practices and results are reported fully and honestly. Large-scale, direct replication by independent researchers is needed to verify the robustness of any promising interventions. Ideally, such replication should occur before policymakers consider recommending an intervention to the public and before the results are used as the basis for marketing claims.

Severe Problems

These problems preclude any conclusions about the causal efficacy of an intervention. Papers with these problems should not be used to draw conclusions about the effectiveness of an intervention. They should not be given media coverage unless the focus of that coverage is on the study's limitations, and they should not be used to guide policy or to promote products.

- No pretest baseline: Without a baseline, differences in outcomes could just reflect differences in pre-training abilities.
- No control group: Without a control group, any improvement observed after an intervention could result from factors other than the intervention itself. Evidence for the effectiveness of an intervention must always be evaluated relative to a baseline.
- Lack of random assignment to conditions: Random assignment is the best way to ensure that uncontrolled factors are equally likely to be true of the intervention and control group. Note, though, that randomization is effective only with relatively large numbers of participants.

Substantial Problems

These problems mean that a study can provide only ambiguous or inconclusive evidence for the effectiveness of an intervention. Findings from papers with these problems should be treated as tentative at most. They should not be used in determining public policy or promoting products without further empirical evidence from more tightly controlled or fully reported studies (ideally, preregistered ones). Media coverage should be cautious and should explicitly note the preliminary and uncertain nature of the evidence.

- Passive control group: Studies comparing an intervention group to a waitlist or no-contact control group cannot attribute causal potency to the intervention itself. Any differences between the treatment and control group can account for the difference (e.g., motivation, expectations, engagement, interaction with the experimenter). Any claim that the intervention itself was effective should be weighed against any of the other factors that might explain the difference. Ideally, studies should include an active control group that is as closely matched to the treatment group as possible, leaving the treatment ingredient itself as the only difference.
- Lack of preregistration: Without preregistration of the testing plan, outcome measures, and analysis plan, interpretation of any reported measures must be tentative. Only with preregistration can readers be certain that a study reported all of the measured outcomes. Without preregistration of testing and analysis plans, the reported results might be cherry-picked (intentionally or unintentionally).
- Scattershot publishing without full documentation: Papers that each report a subset of outcome measures from a study without identifying all of the other outcome measures cannot provide definitive evidence. All statistical tests are suspect because there is no way to correct for the number of tests that could have been conducted.
- Small numbers of participants: With small samples, randomization is ineffective and estimates of the benefits are imprecise. Given that most intervention designs involve an interaction effect—a test of the difference in improvement across groups—they require large samples to have adequate power to detect effects. Samples with less than 10 to 20 participants per group are inadequate; they typically should be an order of magnitude larger when measuring small effects. Small studies are more likely to produce spurious results, a problem that is magnified because spurious results are likely to be published.
- Contingent analyses: Secondary analyses conducted after inspecting the data must be treated as exploratory, even if they are highly significant. They should be confirmed with a preregistered study before they can be considered robust.
- Subgroup analyses: An intervention that unexpectedly works only for one subset of participants and does not work overall could represent a statistical fluke resulting from testing many subgroups; it should be tested again with a preregistered prediction.

Potential Problems

These problems mean that a study might have limitations that were not completely addressed in the published report. Papers with these problems may provide evidence for an intervention's effectiveness, but further analysis or study might undermine their findings. Such papers should be used cautiously in determining public policy and should not be touted in the media or used to promote a product without explicit mention of the study's limitations.

- Active but unmatched control group: Studies that do not match expectations for the critical outcome measures between the intervention group and the active control group cannot provide unambiguous evidence for the effectiveness of the intervention itself. Just because a control group is active does not mean it accounts for differential placebo effects.
- Inadequate preregistration: Vague preregistration plans allow too much flexibility to cherry-pick results (intentionally or unintentionally).
- Departures from preregistration: Studies that do not adhere to their preregistered plans must explicitly acknowledge those departures. If they do not, the results are suspect and should not be trusted.
- Lack of blinding when using subjective outcome measures: Many studies measuring classroom performance or clinical symptoms rely on subjective reports from participants, teachers, parents, or therapists. When the people making such reports are aware of the intervention and condition assignment, they are highly likely to be influenced by expectations. When experimenters are not blind to condition, the use of subjective outcome measures moves into the realm of Substantial Problems.

generalization of training may prove less effective than simply practicing the core real-world skills directly. If your goal is to improve your performance on the trained tasks, then using brain-training software may be an efficient and entertaining way to do so. And such improvements could increase motivation to improve other skills through targeted training. But if your hope is to stave off the cognitive losses that sometimes accompany aging or to enhance your performance at school or in your profession, you should be skeptical of the value of any quick fixes; the evidence largely does not support claims of broad cognitive benefits from practicing the sorts of cognitive tasks used in most brain-training software. Based on the evidence we have reviewed, together with the broader literature on learning and expertise, targeted practice likely will yield greater improvements on the practiced task than will practicing a different task.

Consumers should also consider the comparative costs and benefits of engaging in a brain-training regimen. Time spent using brain-training software could be allocated to other activities or even other forms of "brain training" (e.g., physical exercise) that might have broader benefits for health and well-being. That time might also be spent on learning things that are likely to improve your performance at school (e.g., reading; developing knowledge and skills in math, science, or the arts), on the job (e.g., updating your knowledge of content and standards in your profession), or in activities that are otherwise enjoyable. If an intervention has minimal benefits, using it means you have lost the opportunity to do something else. If you find using brain-training software enjoyable, you should factor that enjoyment into your decision to use it, but you should weigh it against other things you might do instead that would also be enjoyable, beneficial, and/or less expensive.

When evaluating the marketing claims of brain-training companies or media reports of brain-training studies, consider whether they are supported by peer-reviewed, scientific evidence from studies conducted by researchers independent of the company. As we have seen, many brain-training companies cite a large number of papers, but not all of those directly tested the effectiveness of a brain-training program against an appropriate control condition. Moreover, many of the studies tested groups of people who might not be like you. It is not clear that results from studies of people with schizophrenia will generalize to people without schizophrenia, or that benefits found in studies of college students will generalize to older adults. Finally, just because an advertisement appears in a trusted medium (e.g., National Public Radio) or is promoted by a trusted organization (e.g., AARP) does not mean that its claims are justified. Consumers should view such advertising claims with skepticism.

Conclusion

Two consensus statements about brain training, both signed by dozens of scientists, offered conflicting views on the state of the evidence. One argued that no compelling evidence exists to support the claims of brain-training companies that brain games enhance cognition or stave off the cognitive consequences of aging. A rebuttal letter acknowledged that some marketing by brain-training companies has overreached, but cited extensive support for scientifically grounded brain-training interventions.

Based on our extensive review of the literature cited by brain-training companies in support of their claims, coupled with our review of related brain-training literatures that are not currently associated with a company or product, there does not yet appear to be sufficient evidence to justify the claim that brain training is an effective tool for enhancing real-world cognition. The Scottish legal system permits a verdict that eschews "guilty" or "not guilty" in favor of "not proven." Despite marketing claims from brain-training companies of "proven benefits," scientific evidence does not permit claims of "proof." In place of "not proven," we find the evidence of benefits from cognitive brain training to be "inadequate."

Many studies are suggestive of benefits, and the published evidence rarely shows negative consequences of training interventions. Yet most studies have lacked adequate controls for placebo effects, have tested sample sizes that are much too small, and have not fully reported and analyzed all outcome measures. The few large-scale intervention studies provide relatively little support for differential improvements on objective measures of performance in the world, and the results have tended to be distributed across many non-independent papers, making a complete evaluation of the evidence difficult. Methodological standards for intervention research have been relatively lax in much of the brain-training literature, and future research on this important topic should adhere more closely to best practices for intervention research, including preregistration, complete reporting, larger sample sizes, and suitable controls for placebo effects.

Those studies that have reported relatively strong benefits of cognitive training have tended to show relatively narrow transfer—training improved the trained tasks and those that were structurally similar to the trained tasks. Evidence for broader transfer of training is substantially weaker and less common. Few studies provide any evidence that training with brain-training software or basic cognitive tasks yields differential improvements (relative to an appropriate control condition) for cognitive performance in the world. The benefits, when present, have applied almost exclusively to other laboratory tasks.

For our review, we focused on papers cited by brain-training proponents and leading companies as evidence of the effectiveness of cognitive brain training, studies in which people have practiced one or a set of cognitive tasks. Defining our scope in this way means that we did not exhaustively review all published brain-training interventions, including those that involved direct neural stimulation, exercise, nutrition, or other non-cognitive tasks. We hoped to capture those findings that proponents of cognitive brain training believe provide the strongest evidence of benefits. Presumably, companies would not deliberately cite weak evidence if stronger evidence were available. Our limited scope does mean that we might have missed some studies that provide more compelling evidence for transfer of training, but our informal survey of other published research in this literature did not uncover any interventions that better met the best practices we defined.

With a review of this scope, we may well have made some errors in the coding and evaluation of published articles. We do not feel that such errors are likely to affect our global evaluation of the strength of evidence for the effectiveness of brain training. However, we have created and will maintain a list of corrections on the public Open Science Framework page for this project (https://goo.gl/N6jY3s). On that site, we will also identify new, large-scale randomized controlled trials appearing in the scientific literature since the completion of this review. We encourage readers to evaluate new studies in light of the best practices we identify before drawing conclusions about their effectiveness.

Some might argue that our insistence on best practices to justify claims that a product or intervention is effective will stifle innovation or prevent potentially useful products from reaching the market. We disagree. If a company claims scientific proof for the benefits of its products, it must adhere to best scientific practices. Exploratory scientific research is a necessary part of the discovery process, and such research might well lead to innovative and novel approaches to cognitive enhancement. We do not believe our recommendations stifle such discovery. For drug testing, the early phases of research allow for the discovery of promising therapies, but only after more rigorous testing can they be promoted as a treatment. The same is true for cognitive interventions and brain training; an initial discovery using weaker methods should not be translated to the marketplace until it has been evaluated with more rigorous testing.

In sum, despite a large number of published papers reporting tests of the effects of brain-training interventions, the evidence that training with commercial brain-training software can enhance cognition outside the laboratory is limited and inconsistent. The inconsistency of the results and the pervasive shortcomings in research design and analysis in the published literature undermine scientific backing for some of the claims made by brain-training companies. Brain training is appealing in part because it seems to provide a quick way to enhance cognition relative to the sustained investment required by education and skill acquisition. Practicing a cognitive task consistently improves performance on that task and closely related tasks, but the available evidence that such training generalizes to other tasks or to real-world performance is not compelling.

Acknowledgments

We thank an excellent team of student assistants who worked with Simons to gather articles and conduct preliminary screening for relevance: Steven Yu, Marie Bongiorno, Varsha John, Marielle Nagele, Hua Yan, and Emma Roese. Thanks also to Daniel Willingham for helpful suggestions during the early stages of this project.

Author Contributions

All of the authors wrote or coauthored the initial drafts of sections of the manuscript as follows: Daniel J. Simons oversaw the collection of articles from companies and websites and had primary responsibility for coding all of the articles. Simons composed the introduction, the method best practices, the review of evidence from Cognitive Training Data, the review of evidence from brain-training companies other than Cogmed, the recommendations, and the conclusions; integrated all of the sections into the article; and edited them for stylistic consistency. Walter R. Boot composed the sections on video-game interventions, the history of brain training, the review of the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) study, transfer of training, and the review of the Iowa Healthy and Active Minds Study (IHAMS) and the Staying Keen in Later Life (SKILL) study, in addition to parts of other sections. Neil Charness coauthored the review of ACTIVE and sections on comparative costs and interpreting benefits for individuals. Susan E. Gathercole composed the review of Cogmed and background on Cogmed. Christopher F. Chabris wrote on the market for and marketing of brain-training products. David Z. Hambrick composed the review of *n*-back working-memory training. Elizabeth A. L. Stine-Morrow composed the sections on transfer of training, the history of brain training, the review of Posit Science, recommendations for theoretical development and policymakers, and parts of other sections. All authors reviewed sections of the article throughout the process, critically edited the full manuscript draft, and assisted with post-review revisions.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article. Walter R. Boot and Neil Charness are currently conducting an evaluation of an experimental brain-training product with a software package developed by Aptima. They have received no financial support from Aptima, but they are paying Aptima (from a National Institute on Aging grant) to modify its software for the purposes of the study. As a graduate student/postdoc, Boot assisted in the design and implementation of an unpublished intervention study using CogniFit's brain-training software package in collaboration with employees of CogniFit. CogniFit reimbursed Boot for travel to the site of the intervention. Daniel J. Simons contributed to the design of that CogniFit intervention but had no interactions with anyone from CogniFit and received no financial support. Susan E. Gathercole holds a Cogmed research "master license," which provides her with free access to Cogmed training products and support for the purposes of independent research. Christopher F. Chabris is a shareholder and scientific advisor for Knack.it Corporation, a company that develops games to assess cognitive abilities and personality traits. The company's products focus exclusively on assessment. The company does not and never has developed or sold products for the brain-training market as defined in this article. David Z. Hambrick was a co–principal investigator on a grant from the Office of Naval Research (ONR) concerning the effectiveness of cognitive training (ONR Research Grant N000140910129; principal investigator: Randall W. Engle). Elizabeth A. L. Stine-Morrow was a signatory on the Stanford/Max Planck open letter but has no ties to any brain-training companies.

References

"About Lumosity." (2010). [Internet archive]. Retrieved from https://web.archive.org/web/20150905164947/http://www.lumosity.com/about

"About us." (2015, September 5). [Internet archive]. Retrieved from https://web.archive.org/web/20150905140536/http://www.scilearn.com/company/about-us

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2010). Use it or lose it? Wii brain exercise practice and reading for domain knowledge. *Psychology and Aging*, *25*, 753–766.

Adcock, R. A., Dale, C., Fisher, M., Aldebot, S., Genevsky, A., Simpson, G. V., & Vinogradov, S. (2009). When top-down meets bottom-up: Auditory training enhances verbal memory in schizophrenia. *Schizophrenia Bulletin*, *35*, 1132–1141.

Akerlund, E., Esbjörnsson, E., Sunnerhagen, K. S., & Björkdahl, A. (2013). Can computerized working memory Training improve impaired working memory, cognition and psychological health? *Brain Injury*, *27*, 1649–1657.

Albert, M. S., Jones, K., Savage, C. R., Berkman, L., Seeman, T., Blazer, D., & Rowe, J. W. (1995). Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychology and Aging*, *10*, 578–589.

Anders, T. R., & Lillyquist, T. D. (1971). Retrieval time in forward and backward recall. *Psychonomic Science*, *22*, 205–206.

Anderson, J., Reder, L. M., & Simon, H. A. (1996). Situated learning and cognition. *Educational Researcher*, *25*, 5–11.

Anderson, S., White-Schwoch, T., Choi, H. J., & Kraus, N. (2013). Training changes processing of speech cues in older adults with hearing loss. *Frontiers in Systems Neuroscience, 7*, Article 97. doi: 10.3389/fnsys.2013.00097

Anderson, S., White-Schwoch, T., Choi, H. J., & Kraus, N. (2014). Partial maintenance of auditory-based cognitive training benefits in older adults. *Neuropsychologia*, *62*, 286–296.

Angell, J. R., Pillsbury, W. B., & Judd, C. H. (1908). The doctrine of formal discipline in the light of contemporary psychology: A discussion of the proceedings from the meeting of the Michigan Schoolmasters' Club at Ann Arbor, Michigan, April 2, 1908. Retrieved from https://archive.org/details/doctrineofformal00michi.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., . . . Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*, 97–101.

Astle, D. E., Barnes, J. J., Baker, K., Colclough, G. L., & Woolrich, M. W. (2015). Cognitive training enhances intrinsic brain connectivity in childhood. *The Journal of Neuroscience*, *35*, 6277–6283.

Au, J., Buschkuehl, M., Duncan, G. J., & Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychonomic Bulletin & Review*, *23*, 331–337.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*, 366–377.

Avtzon, S. A. (2012). Effect of neuroscience-based cognitive skill training on growth of cognitive deficits associated with learning disabilities in children Grades 2–4. *Learning Disabilities*, *18*, 111–112.

Awh, E., Jonides, J., Smith, E. E., Buxton, R. B., Frank, L. R., Love, T., & Gmeindl, L. (1999). Rehearsal in spatial working memory: Evidence from neuroimaging. *Psychological Science*, *10*, 433–437.

Bagley, W. C. (1909). The psychology of school practice. *Psychological Bulletin*, *6*, 78–83.

Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Journal of the Optical Society of America. A, Optics, Image Science*, *5*, 2210–2219.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, *288*, 2271–2281.

Ball, K., Edwards, J. D., Ross, L. A., & McGwin, G., Jr. (2010). Cognitive training decreases motor vehicle collision involvement among older drivers. *Journal of the American Geriatric Society*, *58*, 2107–2113. doi:10.1111/j.1532-5415.2010.03138.x

Ballesteros, S., Prieto, A., Mayas, J., Toril, P., Pita, C., Ponce de León, L., & Waterworth, J. (2014). Brain training with non-action video games enhances aspects of cognition in older adults: A randomized controlled trial. *Frontiers in Aging Neuroscience*, *6*, Article 277. doi:10.3389/fnagi.2014.00277

Ballesteros, S., Prieto, A., Mayas, J., Toril, P., Pita, C., Ponce de León, L., Reales, J. M., & Waterworth, J. (2015) Corrigendum: Brain training with non-action video games enhances aspects of cognition in older adults: A randomized controlled trial. *Frontiers in Aging Neuroscience*, *7*, Article 82. doi:10.3389/fnagi.2015.00082

Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., DeSouza, S., & Kramer, A. F. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica*, *142*, 74–86.

Barnes, D. E., Yaffe, K., Belfor, N., Jagust, W. J., DeCarli, C., Reed, B. R., & Kramer, J. H. (2009). Computer-based cognitive training for mild cognitive impairment: Results from a pilot randomized, controlled trial. *Alzheimer Disease and Associated Disorders*, *23*, 205.205–10.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637.

Basak, C., Boot, W. R., Voss, M., & Kramer, A. F. (2008). Can training in a real-time strategy videogame attenuate cognitive decline in older adults? *Psychology and Aging*, *23*, 765–777.

Bayarri, M. J., Benjamin, D. J., Berger, J. O., & Sellke, T. M. (2016). Rejection odds and rejection ratios: A proposal for statistical practice in testing hypotheses. *Journal of Mathematical Psychology*, *72*, 90–103.

Beck, S. J., Hanson, C. A., Puffenberger, S. S., Benninger, K. L., & Benninger, W. B. (2010). A controlled trial of working memory training for children and adolescents with ADHD. *Journal of Clinical Child & Adolescent Psychology*, *39*, 825–836.

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z. L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences, USA*, *111*, 16961–16966.

Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association*, *289*, 454–465.

Belchior, P., Marsiske, M., Sisco, S. M., Yam, A., Bavelier, D., Ball, K., & Mann, W. C. (2013). Video game training to improve selective visual attention in older adults. *Computers in Human Behavior*, *29*, 1318–1324.

Bell, M. D., Bryson, G. J., Greig, T. C., Fiszdon, J. M., & Wexler, B. E. (2005). Neurocognitive enhancement therapy with work therapy: Productivity outcomes at 6-and 12-month follow-ups. *Journal of Rehabilitation Research & Development*, *42*, 829–838.

Bell, M. D., Zito, W., Greig, T., & Wexler, B. E. (2008a). Neurocognitive enhancement therapy and competitive employment in schizophrenia: Effects on clients with poor community functioning. *American Journal of Psychiatric Rehabilitation*, *11*, 109–122.

Bell, M. D., Zito, W., Greig, T., & Wexler, B. E. (2008b). Neurocognitive enhancement therapy with vocational services: Work outcomes at two-year follow-up. *Schizophrenia Research*, *105*, 18–29.

Bell, M., Bryson, G., Greig, T., Corcoran, C., & Wexler, B. E. (2001). Neurocognitive enhancement therapy with work therapy: Effects on neuropsychological test performance. *Archives of General Psychiatry*, *58*, 763–768.

Bell, M., Bryson, G., & Wexler, B. E. (2003). Cognitive remediation of working memory deficits: Durability of training effects in severely impaired and less severely impaired schizophrenia. *Acta Psychiatrica Scandinavica*, *108*, 101–109.

Bell, M., Fiszdon, J., Greig, T., Wexler, B., & Bryson, G. (2007). Neurocognitive enhancement therapy with work therapy in schizophrenia: 6-month follow-up of neuropsychological performance. *Journal of Rehabilitation Research and Development*, *44*, 761–770.

Bennett, C. J. C. (1907). *Formal discipline*. New York, NY: Teachers College, Columbia University.

Bennett, S., Holmes, J., & Buckley, S. (2013). Computerized memory training leads to sustained improvement in visuospatial short term memory skills in children with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *118*, 179–192. doi:10.1352/1944-7558-118.3.179

Bentwich, J., Dobronevsky, E., Aichenbaum, S., Shorer, R., Peretz, R., Khaigrekht, M., . . . Rabey, J. M. (2011). Beneficial effect of repetitive transcranial magnetic stimulation combined with cognitive training for the treatment of Alzheimer's disease: A proof of concept study. *Journal of Neural Transmission*, *118*, 463–471.

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*, 591–601.

Berry, A. S., Zanto, T. P., Clapp, W. C., Hardy, J. L., Delahunt, P. B., Mahncke, H. W., & Gazzaley, A. (2010). The influence of perceptual training on working memory in older adults. *PLoS ONE*, *5*(7), e11537. doi:10.1371/journal.pone.0011537

Best Funny Commercial. (2015, June 4). Inside Out *Lumosity commercial 2015 HD* [Video file]. Retrieved from https://www.youtube.com/watch?v=2oM1YV21Wlg

Blacker, K. J., Curby, K. M., Klobusicky, E., & Chein, J. M. (2014). Effects of action video game training on visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1992–2004.

Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology*, *2*, 247–265.

Bloomberg. (2016, August 1). Company overview of Posit Science Corporation. Retrieved from http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=7757398

Boot, W. R., Champion, M., Blakely, D. P., Wright, T., Souders, D. J., & Charness, N. (2013). Video games as a means to reduce age-related cognitive decline: Attitudes, compliance, and effectiveness. *Frontiers in Psychology, 4*, Article 31. doi:10.3389/fpsyg.2013.00031

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*, 387–398.

Boot, W. R., Simons, D. J., 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.

Bosma, H., van Boxtel, M. P., Ponds, R. W. H. M., Jelicic, M., Houx, P., Metsemakers, J., & Jolles, J. (2002). Engaged lifestyle and cognitive function in middle and old-aged, non-demented persons: A reciprocal association? *Zeitschrift für Gerontologie und Geriatrie*, *35*, 575–581.

Bouchard, C., Sarzynski, M. A., Rice, T. K., Kraus, W. E., Church, T. S., Sung, Y. J., . . . Rankinen, T. (2011). Genomic predictors of maximal O2 uptake response to standardized exercise training programs. *Journal of Applied Physiology*, *110*, 1160–1170. doi:10.1152/japplphysiol.00973.2010

Boudreaux, D. J. (1995, September). "Puffery" in advertising. *The Free Market*, *13*(9). Retrieved from https://mises.org/library/puffery-advertising

Bowie, C. R., Gupta, M., Holshausen, K., Jokic, R., Best, M., & Milev, R. (2013). Cognitive remediation for treatment-resistant depression: Effects on cognition and functioning and the role of online homework. *The Journal of Nervous and Mental Disease*, *201*, 680–685.

Bowie, C. R., McGurk, S. R., Mausbach, B., Patterson, T. L., & Harvey, P. D. (2014). Combined cognitive remediation and functional skills training for schizophrenia: Effects on cognition, functional competence, and real-world behavior. *American Journal of Psychiatry*, *169*, 710–718.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, *301*, 419–421.

"Brain Age: Concentration training." (2012). Retrieved from http://brainage.nintendo.com/why-concentration-training/

"Brain Age: Devilish training." (2012). Retrieved from http://brainage.nintendo.com/devilish-training/

"Brain Age on TV." (2012, January 16). [Internet archive]. Retrieved from https://web.archive.org/web/20120116164307/http://brainage.com/launch/ontv.jsp?video=tvspot

"Brain training that works." (2015, September 9). [Internet archive]. Retrieved from https://web.archive.org/web/20150909153709/http://www.brainhq.com

"Brain training that works." (2016). Retrieved from http://easterseals.brainhq.com/

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, *6*, Article 63. doi:10.3389/fnhum.2012.00063

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., . . . Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences, USA*, *107*, 7939–7944.

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.

Burkhalter, T. M., & Hillman, C. H. (2011). A narrative review of physical activity, nutrition, and obesity to cognition and scholastic performance across the human lifespan. *Advances in Nutrition*, *2*, 201S–206S.

Burman, R. (2007, June 25). Nicole Kidman trains her brain. *IGN.com*. Retrieved from http://www.ign.com/articles/2007/06/25/nicole-kidman-trains-her-brain

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P., Sims, B. C., Jewkes, A. M., & Morrison, F. J. (2008). Age- and schooling related effects on executive functions in young children: A natural experiment. *Child Neuropsychology*, *14*, 510–524.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., . . . Ramon, M. (2014). A randomized clinical trial of Cogmed working memory training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, *55*, 247–255.

Chase, W. G., & Ericsson, K. A. (1982). Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1–58). New York, NY: Academic Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Cherney, I. D. (2008). Mom, let me play more computer games: They improve my mental rotation skills. *Sex Roles*, *59*, 776–786.

Cherney, I. D., Bersted, K., & Smetter, J. (2014). Training spatial skills in men and women. *Perceptual and Motor Skills*, *119*, 82–99.

Chooi, W., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*, 531–542. doi:10.1016/j.intell.2012.07.004

Clark, J. E., Lanphear, A. K., & Riddick, C. C. (1987). The effects of videogame playing on the response selection processing of elderly adults. *Journal of Gerontology*, *42*, 82–85. doi:10.1093/geronj/42.1.82

Clay, O. J., Wadley, V. G., Edwards, J. D., Roth, D. L., Roenker, D. L., & Ball, K. K. (2005). Cumulative meta-analysis of the relationship between useful field of view and driving performance in older adults: Current and future implications. *Optometry and Vision Science*, *82*, 724–731.

Cogan, A., Madey, J., Kaufman, W., Holmlund, G., & Bach-y-Rita, P. (1977). Pong game as a rehabilitation device. In G. C. Warren (Ed.), *Fourth Annual Conference on Systems and Devices for the Disabled* (pp. 187–188). Seattle: University of Washington School of Medicine.

"Cogmed solutions help consumers, professionals, and schools address attention problems." (2015). Retrieved from http://www.cogmed.com/program

"A consensus on the brain training industry from the scientific community." (2014). Retrieved from http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community-2/

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W. . . . European ADHD Guidelines Group. (2015). Cognitive training for Attention-Deficit/Hyperactivity Disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*, 164–174.

"Cutting edge neuroscience personalized for you." (2015, September 15). [Internet archive]. Retrieved from https://web.archive.org/web/20150915095931/http://www.lumosity.com/

"Cutting edge science personalized for you." (2015, September 4). [Internet archive]. Retrieved from https://web.archive.org/web/20150904041409/http://www.lumosity.com

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*, 1510–1512.

David, P., & Gelfeld, V. (2014, September 17). *Brain Health Research Study*. AARP. Retrieved from http://www.aarp.org/content/dam/aarp/research/surveys_statistics/health/2015/2014-Brain-Health-Research-Study-AARP-res-gen.pdf

David, P., & Gelfeld, V. (2015, January 20). 2014 Brain Health Research Study. Retrieved from http://www.aarp.org/research/topics/health/info-2015/staying-sharper-study.html

Deary, I. J. (2008). Why do intelligent people live longer? *Nature*, *13*, 175–176.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453–482.

de Groot, A. D. (1978). *Het denken van de schaker* [The thought of the chess player]. Amsterdam, The Netherlands: Mouton. (Original work published 1946)

DeGutis, J. M., & Van Vleet, T. M. (2010). Tonic and phasic alertness training: A novel behavioral therapy to improve spatial and non-spatial attention in patients with hemispatial neglect. *Frontiers in Human Neuroscience, 4*, Article 60. doi:10.3389/fnhum.2010.00060

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology*, *163*, 272–282.

Dennis, K. A., & Harris, D. (1998). Computer based simulation as an adjunct to Ab Initio flight training. *The International Journal of Aviation Psychology*, *8*, 261–277.

DiMauro, J., Genova, M., Tolin, D. F., & Kurtz, M. M. (2014). Cognitive remediation for neuropsychological impairment in hoarding disorder: A pilot study. *Journal of Obsessive-Compulsive and Related Disorders*, *3*, 132–138.

Donchin, E., Fabiani, M., & Sanders, A. (Eds.). (1989). The learning strategies program: An examination of the strategies in skill acquisition [Special issue]. *Acta Psychologica*, *71*, 1–309.

Dorval, M., & Pepin, M. (1986). Effect of playing a video game on a measure of spatial visualization. *Perceptual and Motor Skills*, *62*, 159–162.

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin & Review*, *23*, 206–316.

Dovis, S., van Rentergem, J. A., & Huizenga, H. M. (2015, April 4). Does Cogmed working memory training really improve inattention in daily life? A reanalysis [Reader comment]. Retrieved from http://journals.plos.org/plosone/article/comment?id=info:doi/10.1371/annotation/148c2260-54ce-4d85-a49b-740053f6dbc8

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *The Journal of Neuroscience*, *26*, 6314–6317.

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioural Brain Research*, *192*, 137–142.

"Drivesharp online training." (2015, May 2). [Internet archive]. Retrieved from https://web.archive.org/web/20150502194800/https://www.usaa.com/inet/pages/drivesharp_online_product_page_complex_module

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58.

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition*, *42*, 854–862.

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, *16*, 915–925. doi:10.1111/desc.12068

Edwards, J. D., Delahunt, P. B., & Mahncke, H. W. (2009). Cognitive speed of processing training delays driving cessation. *The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences*, *64*, 1262–1267.

Edwards, J. D., Hauser, R. A., O'Connor, M. L., Valdés, E. G., Zesiewicz, T. A., & Uc, E. Y. (2013). Randomized trial of cognitive speed of processing training in Parkinson disease. *Neurology*, *81*, 1284–1290.

Edwards, J. D., Myers, C., Ross, L. A., Roenker, D. L., Cissell, G. M., McLaughlin, A. M., & Ball, K. K. (2009). The longitudinal impact of cognitive speed of processing training on driving mobility. *The Gerontologist*, *49*, 485–494.

Edwards, J. D., Valdés, E. G., Peronto, C., Castora-Binkley, M., Alwerdt, J., Andel, R., & Lister, J. J. (2013). The efficacy of InSight cognitive training to improve useful field of view performance: A brief report. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences*, *70*, 417–422.

Edwards, J. D., Vance, D. E., Wadley, V. G., Cissell, G. M., Roenker, D. L., & Ball, K. K. (2005). Reliability and validity of the Useful Field of View test scores as administered by personal computer. *Journal of Clinical and Experimental Neuropsychology*, *27*, 529–543.

Edwards, J. D., Wadley, V. G., Meyers, R. S., Roenker, D. R., Cissell, G. M., & Ball, K. K. (2002). Transfer of a speed of processing intervention to near and far cognitive functions. *Gerontology*, *48*, 329–340.

Egeland, J., Aarlien, A. K., & Saunes, B.-K. (2013). Few effects of far transfer of working memory training in ADHD: A randomized controlled trial. *PLoS ONE*, *8*(10), e75660. doi:10.1371/journal.pone.0075660

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*, 3–6.

Ericsson, K. A. (2006). The influence in experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge, England: Cambridge University Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.

Federal Trade Commission. (2016a). [Press release]. Retrieved from https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges

Federal Trade Commission. (2016b). *Complaint against Lumos Labs*. Retrieved from https://www.ftc.gov/system/files/documents/cases/160105lumoslabscmpt.pdf

Federal Trade Commission. (2016c). *Settlement with Lumos Labs*. Retrieved from https://www.ftc.gov/system/files/documents/cases/160105lumoslabsstip.pdf

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850–855.

Fernandez, A. (2009, July 14). AAA to deploy Brain Fitness Software DriveSharp to assess and train older driver's brains [Blog post]. Retrieved from http://sharpbrains.com/blog/2009/07/14/aaa-to-deploy-brain-fitness-software-drivesharp-to-assess-and-train-older-drivers-brains/

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., . . . Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, *25*, 736–744. doi:10.1177/0956797613516008

Finn, M., & McDonald, S. (2011). Computerised cognitive training for older persons with mild cognitive impairment: A pilot study using a randomised controlled trial design. *Brain Impairment*, *12*, 187–199.

Fisher, M., Holland, C., Merzenich, M. M., & Vinogradov, S. (2009). Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. *The American Journal of Psychiatry*, *166*, 805–811.

Fisher, M., Holland, C., Subramaniam, K., & Vinogradov, S. (2010). Neuroplasticity-based cognitive training in schizophrenia: An interim report on the effects 6 months later. *Schizophrenia Bulletin*, *36*, 869–879.

Fisher, M., Loewy, R., Carter, C., Lee, A., Ragland, J. D., Niendam, T., . . . Vinogradov, S. (2015). Neuroplasticity-based auditory training via laptop computer improves cognition in young individuals with recent onset schizophrenia. *Schizophrenia Bulletin*, *41*, 250–258.

Fisher, M., Loewy, R., Hardy, K., Schlosser, D., & Vinogradov, S. (2013). Cognitive interventions targeting brain plasticity in the prodromal and early phases of schizophrenia. *Annual Review of Clinical Psychology*, *9*, 435–463.

Fiszdon, J. M., Bryson, G. J., Wexler, B. E., & Bell, M. D. (2004). Durability of cognitive remediation training in schizophrenia: Performance on two memory tasks at 6-month and 12-month follow-up. *Psychiatry Research*, *125*, 1–7.

Foy, J. G., & Mann, V. A. (2014). Adaptive cognitive training enhances executive control and visuospatial and verbal working memory in beginning readers. *International Education Research*, *2*(2), 19–43.

Franck, N., Duboc, C., Sundby, C., Amado, I., Wykes, T., Demily, C., . . . Cognitive Remediation Network. (2013). Specific vs general cognitive remediation for executive functioning in schizophrenia: A multicenter randomized trial. *Schizophrenia Research*, *147*, 68–74.

"Frequently asked questions." (2015). Retrieved from http://www.cogmed.com/faq

"Frequently asked questions." (n.d.). Retrieved from http://ies.ed.gov/ncee/wwchttp://ies.ed.gov/ncee/wwc/document.aspx?sid=15

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *Educational Communications and Technology Journal*, *33*, 263–275.

Gard, T., Hölzel, B. K., & Lazar, S. W. (2014). The potential effects of meditation on age-related cognitive decline: A systematic review. *Annals of the New York Academy of Sciences*, *1307*, 89–103.

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, P., Devereaux, P. J., Beyene, J., . . . Haynes, B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, *293*, 1223–1238.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*, 1–16.

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem even when there is no "fishing expectation" or "p-hacking" and the research hypothesis was posited ahead of time*. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 305–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Gillam, R. B., Frome Loeb, D., Hoffman, L. M., Bohman, T., Champlin, C. A., & Thibodeau, L. (2008). The Efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, *51*, 97–119.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*, 159–163.

Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., . . . Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: A randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *53*, 1277–1284.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.

Green, C. S., & Bavelier, D. (2006a). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1465–1468.

Green, C. S., & Bavelier, D. (2006b). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*, 217–245.

Green, C. S., & Bavelier, D. (2007). Action video game experience alters the spatial resolution of attention. *Psychological Science*, *18*, 88–94.

Green, C. S., Pouget, A., & Bavelier, D. (2010). A general mechanism for learning with action video games: Improved probabilistic inference. *Current Biology*, *20*, 1573–15792.

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*, 756–772.

Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Bavelier, D. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior*, *28*, 984–994.

Green, C. T., Long, D. L., Green, D., Iosif, A., Dixon, F., Miller, M. R., . . . Schweitzer, J. B. (2012). Will working memory training generalize to improve off-task behavior in children with Attention-Deficit/Hyperactivity Disorder? *Neurotherapeutics*, *9*, 639–648.

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H., & Kaye, D. (1994). Action video games and informal education: Effects on strategies for dividing visual attention. *Journal of Applied Developmental Psychology*, *15*, 105–123.

Greig, T. C., Zito, W., Wexler, B. E., Fiszdon, J., & Bell, M. D. (2007). Improved cognitive function in schizophrenia after one year of cognitive training and vocational services. *Schizophrenia Research*, *96*, 156–161.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills*, *57*, 155–158.

Gropper, R. J., Gotlieb, H., Kronitz, R., & Tannock, R. (2014). Working memory training in college students with ADHD or LD. *Journal of Attention Disorders*, *18*, 331–345.

Grossmann, I., Karasawa, M., Izumi, S., Na, J., Varnum, M. E. W., Kitayama, S., & Nisbett, R. (2012). Age and wisdom: Culture matters. *Psychological Science*, *23*, 1059–1066.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.

Habib, M., Rey, V., Daffaure, V., Camps, R., Espesser, R., Joly-Pottuz, B., & Démonet, J. F. (2002). Phonological training in children with dyslexia using temporally modified speech: A three-step pilot investigation. *International Journal of Language & Communication Disorders*, *37*, 289–308.

Haimov, I., & Shatil, E. (2013). Cognitive training improves sleep quality and cognitive function among older adults with insomnia. *PLoS ONE*, *8*(4), e61390. doi:10.1371/journal.pone.0061390

Hale, S., Myerson, J., Faust, M., & Fristoe, N. (1995). Converging evidence for domain-specific slowing from multiple non-lexical tasks and multiple analytic methods. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences, 50B*, P202–P211.

Hardy, J. L., Drescher, D., Sarkar, K., Kellett, G., & Scanlon, M. (2011). Enhancing visual attention and working memory with a web-based cognitive training program. *Mensa Research Journal*, *42*(2), 13–20.

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. *PLoS ONE*, *10*(9), e0134467. doi:10.1371/journal.pone.0134467

Hardy, K. K., Willard, V. W., Allen, T. M., & Bonner, M. J. (2013). Working memory training in survivors of pediatric cancer: A randomized pilot study. *Psycho-Oncology*, *22*, 1856–1865. doi:10.1002/pon.3222

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*, 2409–2419.

Hay, M., Rosenthal, J., Thomas, D., & Craighead, J. (2011). "BIO / BioMedTracker clinical trial success rates study" [Handout]. Retrieved from http://www.biotech-now.org/wp-content/uploads/2011/02/bio-ceo-biomedtracker-bio-study-handout-final-2-15-2011.pdf

Healy, A. F., & Bourne, L. E. (Eds.). (2012). *Training cognition: Optimizing efficiency, durability, and generalizability*. New York, NY: Psychology Press.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*, 486–502.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest*, *9*, 1–65.

Hewins, N. P. (1916). *The doctrine of formal discipline in the light of experimental investigation*. Baltimore, MD: Warwick & York.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nature Reviews Neuroscience*, *9*, 59–65.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15.

Holmes, J., Hilton, K., Place, M., Alloway, T. P., Elliott, J., & Gathercole, S. E. (2014). Children with low working memory and children with ADHD: Same or different? *Frontiers in Human Neuroscience*, *8*, Article 976. doi:10.3389/fnhum.2014.00976

Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science*, *6*, 537–559.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind–brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience*, *25*, 1397–1405.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties—A longitudinal study. *Annals of Dyslexia*, *51*, 75–96.

Hooker, C. I., Bruce, L., Fisher, M., Verosky, S. C., Miyakawa, A., D'Esposito, M., & Vinogradov, S. (2013). The influence of combined cognitive plus social-cognitive training on amygdala response during face emotion recognition in schizophrenia. *Psychiatry Research: Neuroimaging*, *213*, 99–107.

Hooker, C. I., Bruce, L., Fisher, M., Verosky, S. C., Miyakawa, A., & Vinogradov, S. (2012). Neural activity during emotion recognition after combined cognitive plus social cognitive training in schizophrenia. *Schizophrenia Research*, *139*, 53–59.

"How is Cogmed different." (2015). Retrieved from http://www.cogmed.com/how-is-cogmed-different

"How is Cogmed different from other cognitive training programs?" (2015). Retrieved from http://www.cogmed.com/cogmed-cognitive-training-programs

"The Human Cognition Project: Bibliography." (n.d.). Retrieved from http://www.lumosity.com/hcp/research/bibliography

Hurley, D. (2013). *Smarter: The new science of building brain power*. New York, NY: Avery.

"Improve your brain while having fun." (2015, September 6). [Internet archive]. Retrieved from https://web.archive.org/web/20150906125111/https://www.cognifit.com/

Ioannidis, J. P. (2005). Why most published research findings are false. *Chance*, *18*(4), 40–47.

Jacoby, N., & Ahissar, M. (2015). Assessing the applied benefits of perceptual training: Lessons from studies of training working-memory. *Journal of Vision*, *15*, Article 6. doi:10.1167/15.10.6

Jaeggi, S. M. (2005). *Capacity limitations in human cognition: Behavioural and biological considerations* (Unpublished doctoral dissertation). University of Bern, Switzerland.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, USA*, *105*, 6829–6833. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences, USA*, *108*, 10081–10086.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394–412.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*, 464–480.

James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt and Co.

Japancommercials4U2. (2009, October 26). *Brain Age: Train your brain in minutes a day* [Video file]. Retrieved from https://www.youtube.com/watch?v=WlfjHCEWpXY

Jastrzembski, T. S., & Charness, N. (2007). The Model Human Processor and the older adult: Parameter estimation and validation within a mobile phone task. *Journal of Experimental Psychology: Applied*, *13*, 224–248.

Jobe, J. B., Smith, D. M., Ball, K., Tennstedt, S. L., Marsiske, M., Willis, S. L., . . . Kleinman, K. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials*, *22*, 453–479.

Johnston, R. S., Johnson, C., & Gray, C. (1987). The emergence of the word length effect in young children: The effects of overt and covert rehearsal. *British Journal of Developmental Psychology*, *5*, 243-248.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 615–622.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47–70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE*, *10*(8), E0132382. doi:10.1371/journal.pone.0132382

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*, 2027–2037.

Kaufman, S. B. (2014, January 27). *Can cognitive training make you smarter? Interview with author Dan Hurley.* Scientific American Blogs. Retrieved from http://blogs.scientificamerican.com/beautiful-minds/can-brain-training-make-you-smarter-interview-with-author-dan-hurley/

Keefe, R. S., Vinogradov, S., Medalia, A., Buckley, P. F., Caroff, S. N., D'Souza, D. C., . . . Stroup, T. S. (2012). Feasibility and pilot efficacy results from the multi-site Cognitive Remediation in the Schizophrenia Trials Network (CRSTN) study. *The Journal of Clinical Psychiatry*, *73*, 1016–1022.

Kesler, S., Hadi Hosseini, S. M., Heckler, C., Janelsins, M., Palesh, O., Mustian, K., & Morrow, G. (2013). Cognitive training for improving executive function in chemotherapy-treated breast cancer survivors. *Clinical Breast Cancer*, *13*, 299–306. doi:10.1016/j.clbc.2013.02.004

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 177–186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*, 781–791.

Knapp, M., Thorgrimsen, L., Patel, A., Spector, A., Hallam, A., Woods, B., & Orrell, M. (2006). Cognitive stimulation therapy for people with dementia: Cost-effectiveness analysis. *The British Journal of Psychiatry*, *188*, 574–580.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: A comparison of young and old adults. *Journal of Experimental Psychology: Applied*, *1*, 50–76.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Kurtz, M. M., Seltzer, J. C., Fujimoto, M., Shagan, D. S., & Wexler, B. E. (2009). Predictors of change in life skills in schizophrenia after cognitive remediation. *Schizophrenia Research*, *107*, 267–274.

Kurtz, M. M., Seltzer, J. C., Shagan, D. S., Thime, W. R., & Wexler, B. E. (2007). Computer-assisted cognitive remediation in schizophrenia: What is the active ingredient? *Schizophrenia Research*, *89*, 251–260.

Kurtz, M. M., Wexler, B. E., Fujimoto, M., Shagan, D. S., & Seltzer, J. C. (2008). Symptoms versus neurocognition as predictors of change in life skills in schizophrenia after outpatient rehabilitation. *Schizophrenia Research*, *102*, 303–311.

Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*, 278–292.

Lampit, A., Hallock, H., Moss, R., Kwok, S., Rosser, M., Lukjanenko, M., . . . Valenzuela, M. (2014). The time-course of global cognitive gains from supervised computer-assisted cognitive training: A randomised, active-controlled trial in elderly with multiple dementia risk factors. *Journal of Prevention of Alzheimer's Disease*, *1*, 33–39.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, *43*, 431–442.

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, *26*, 952–960.

Lenox, R. H., McNamara, R. K., Papke, R. L., & Manji, H. K. (1999). Neuobiology of lithium: An update. *The Journal of Clinical Psychiatry*, *59*(Suppl. 6), 37–47.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*, 626–629.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*, 549–551.

Li, R., Polat, U., Scalzo, F., & Bavelier, D. (2010). Reducing backward masking through action game training. *Journal of Vision*, *10*(14), Article 33. doi:10.1167/10.14.33

Loewenstein, D. A., Acevedo, A., Czaja, S. J., & Duara, R. (2004). Cognitive rehabilitation of mildly impaired Alzheimer disease patients on cholinesterase inhibitors. *The American Journal of Geriatric Psychiatry*, *12*, 395–402.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaeffer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*, 659–676.

"Lumosity to pay $2M over allegations that its ads deceived customers." [Video file]. (2016, January 5). *NBC Nightly News*. Retrieved from http://www.nbcnews.com/nightly-news/video/lumosity-to-pay--2m-over-allegations-that-its-ads-deceived-customers-596362307693

Lundqvist, A., Grundström, K., Samuelsson, K., & Rönnberg, J. (2010). Computerized training of working memory in a group of patients suffering from acquired brain injury. *Brain Injury*, *24*, 1173–1183.

Luppa, M., Luck, T., Weyrer, S., König, H.-M., Brähler, E., & Riedel-Heller, S. G. (2010). Prediction of institutionalization in the elderly. A systematic review. *Age and Ageing*, *39*, 31–38. doi:10.1093/ageing/afp202

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences, USA*, *97*, 4398–4403.

Maguire, E. A., Spiers, H. J., Good, C. D., Hartley, T., Frakowiak, R. S. J., & Burgess, N. (2003). Navigation expertise and the human hippocampus: A structural brain imaging analysis. *Hippocampus*, *13*, 208–217.

Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Brain plasticity and functional losses in the aged: Scientific bases for a novel intervention. *Progress in Brain Research*, *157*, 81–109.

Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., . . . Merzenich, M. M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences, USA*, *103*, 12523–12528.

Marler, J. A., Champlin, C. A., & Gillam, R. B. (2001). Backward and simultaneous masking measured in children with language-learning impairments who received intervention with Fast ForWord or Laureate Learning Systems software. *American Journal of Speech-Language Pathology*, *10*, 258–268.

Mayas, J., Parmentier, F. B. R., Andres, P., & Ballesteros, S. (2014). Plasticity of attentional functions in older adults after non-action video game training: A randomized, controlled trial. *PLoS ONE*, *9*(3), e92269. doi:10.1371/journal.pone.0092269

Mazer, B. L., Sofer, S., Korner-Bitensky, N., Gelinas, I., Hanley, J., & Wood-Dauphinee, S. (2003). Effectiveness of a visual attention retraining program on the driving performance of clients with stroke. *Archives of Physical Medicine and Rehabilitation*, *84*, 541–550.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343–352.

McDougall, S., & House, B. (2012). Brain training in older adults: Evidence of transfer to memory span performance and pseudo-Matthew effects. *Aging, Neuropsychology, and Cognition*, *19*, 195–221.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*, 10–17.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291.

Melby-Lervåg, M., Redick, T., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*, 512–534.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non) persuasive power of a brain image. *Psychonomic Bulletin & Review*, *20*, 720–725.

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology*, *42*, 850–864.

Miller, K. J., Dye, R. V., Kim, J., Jennings, J. L., O'Toole, E., Wong, J., & Siddarth, P. (2013). Effect of a computerized brain exercise program on cognitive performance in older adults. *The American Journal of Geriatric Psychiatry*, *21*, 655–663.

Mishra, J., Rolle, C., & Gazzaley, A. (2014). Neural plasticity underlying visual perceptual learning in aging. *Brain Research*, *1612*, 140–151.

Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association*, *285*, 1987–1991.

Moreau, D. (2013). Differentiating two-from three-dimensional mental rotation training effects. *The Quarterly Journal of Experimental Psychology*, *66*, 1399–1413.

Morimoto, S. S., Wexler, B. E., Liu, J., Hu, W., Seirup, J., & Alexopoulos, G. S. (2014). Neuroplasticity-based computerized cognitive remediation for treatment-resistant geriatric depression. *International Journal of Geriatric Psychiatry,* 27, 1239–1247.

National Institutes of Health. (2010). NIH Consensus Development Conference statement on preventing Alzheimer's disease and cognitive decline. Retrieved from https://consensus.nih.gov/2010/docs/alz/ALZ_Final_Statement.pdf

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Nintendo Corp. (2015). *Top selling software units: Nintendo DS*. Retrieved from http://www.nintendo.co.jp/ir/en/sales/software/ds.html

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, *238*, 625–631.

Noack, H., Lövdén, M., Schmiedek, F., & Linderberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience*, *27*, 435–453.

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., . . . Kawashima, R. (2012). Brain training game improves executive functions and processing speed in the elderly: A randomized controlled trial. *PLoS ONE*, *7*(1), e29676. doi:10.1371/journal.pone.0029676

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Nozawa, T., Kambara, T., . . . Kawashima, R. (2013). Brain training game boosts executive functions, working memory and processing speed in the young adults: A randomized controlled trial. *PLoS ONE*, *8*(2), e55518. doi:10.1371/journal.pone.0055518

O'Brien, J. L., Edwards, J. D., Maxfield, N. D., Peronto, C. L., Williams, V. A., & Lister, J. J. (2013). Cognitive training and selective attention in the aging brain: An electrophysiological study. *Clinical Neurophysiology*, *124*, 2198–2208.

Oei, A. C., & Patterson, M. D. (2013). Enhancing cognition with video games: A multiple game training study. *PLoS ONE*, *8*(3), e58546. doi:10.1371/journal.pone.0058546

Oei, A. C., & Patterson, M. D. (2014). Playing a puzzle video game with changing requirements improves executive functions. *Computers in Human Behavior*, *37*, 216–228.

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15*, 33–58.

Owsley, C., Ball, K., McGwin, G., Jr., Sloane, M. E., Roenker, D. L., White, M. F., & Overley, E. T. (1998). Visual processing impairment and risk of motor vehicle crash among older adults. *Journal of the American Medical Association*, *279*, 1083–1088. doi:10.1001/jama.279.14.1083

Owsley, C., Ball, K., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1991). Visual/cognitive correlates of vehicle accidents in older drivers. *Psychology and Aging*, *6*, 403–415.

Peretz, C., Korczyn, A. D., Shatil, E., Aharonson, V., Birnboim, S., & Giladi, N. (2011). Computer-based, personalized cognitive training versus classical computer games: A randomized double-blind prospective trial of cognitive stimulation. *Neuroepidemiology*, *36*, 91–99.

Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *The American Journal of Psychiatry*, *162*, 1957–1960.

"Philips brain fitness powered by BrainHQ." (2015, September 10). [Internet archive]. Retrieved from https://web.archive.org/web/20150910052723/http://www.lifelinesys.com/BrainFitness

Pimperton, H., & Nation, K. (2012). Poor comprehenders in the classroom: Teacher ratings of behavioural outcomes in children with specific reading comprehension difficulties. *Journal of Learning Disabilities*, *47*, 199–207.

Plemons, J. K., Willis, S. L., & Baltes, P. B. (1978). Modifiability of fluid intelligence in aging: A short-term longitudinal training approach. *Journal of Gerontology*, *22*, 224–231.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics* (5th ed.). New York, NY: Worth Publishers.

Polat, U., Schor, C., Tong, J. L., Zomet, A., Lev, M., Yehezkel, O., . . . Levi, D. M. (2012). Training the brain to overcome the effect of aging on the human eye. *Scientific Reports*, *2*, 278.

Popov, T., Jordanov, T., Rockstroh, B., Elbert, T., Merzenich, M. M., & Miller, G. A. (2011). Specific cognitive training normalizes auditory sensory gating in schizophrenia: A randomized trial. *Biological Psychiatry*, *69*, 465–471.

Popov, T., Rockstroh, B., Weisz, N., Elbert, T., & Miller, G. A. (2012). Adjusting brain dynamics in schizophrenia by means of perceptual and cognitive training. *PLoS ONE*, *7*(7), e39051. doi:10.1371/journal.pone.0039051

Posit Science Corporation. (2010). DriveSharp: Overview + installation [Manual]. Retrieved from http://www.brainhq.com/sites/default/files/pdfs/ds_install.pdf

Postman, L. (1972). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Learning, Motivation, and Memory: Vol. II. Woodworth & Schlosberg's experimental psychology* (3rd ed., pp. 1019–1132). New York, NY: Holt, Rinehart and Winston.

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*, *20*, 1055–1079.

Preiss, M., Shatil, E., Čermáková, R., Cimermanová, D., & Ram, I. (2013). Personalized cognitive training in unipolar and bipolar disorder: A study of cognitive functioning. *Frontiers in Human Neuroscience*, *7*, Article 108. doi:10.3389/fnhum.2013.00108

Pressler, S. J., Therrien, B., Riley, P. L., Chou, C. C., Ronis, D. L., Koelling, T. M., . . . Giordani, B. (2011). Nurse-enhanced memory intervention in heart failure: The MEMOIR study. *Journal of Cardiac Failure*, *17*, 832–843.

Rabey, J. M., Dobronevsky, E., Aichenbaum, S., Gonen, O., Marton, R. G., & Khaigrekht, M. (2013). Repetitive transcranial magnetic stimulation combined with cognitive training is a safe and effective modality for the treatment of

Alzheimer's disease: A randomized, double-blind study. *Journal of Neural Transmission*, *120*, 813–819.

Rabipour, S., & Davidson, P. S. (2015). Do you believe in brain training? A questionnaire about expectations of computerised cognitive training. *Behavioural Brain Research*, *295*, 64–70.

Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, *33*, 1237–1252.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*, 31–74.

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., . . . Willis, S. L. (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*, 16–24.

Redick, T. S. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence*, *50*, 14–20.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379.

Redick, T. S., & Webster, S. B. (2014). Videogame interventions and spatial ability interactions. *Frontiers in Human Neuroscience*, *8*, Article 183. doi:10.3389/fnhum.2014.00183

Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, *28*, 515–521.

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability or in specific skills? *Developmental Psychology*, *51*, 573–582.

Roberts, G., Quach, J., Gold, L., Anderson, P., Rickards, F., Mensah, F., . . . Wake, M. (2011). Can improving working memory prevent academic difficulties? A school based randomised controlled trial. *BMC Pediatrics, 11*, Article 57. doi:10.1186/1471-2431-11-57

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P., Gathercole, S., Gold, L., . . . Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *JAMA Pediatrics*, *170*, e154568. doi:10.1001/jamapediatrics.2015.4568

Roediger, H. L. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest*, *14*, 1–3.

Roenker, D. L., Cissell, G. M., Ball, K. K., Wadley, V. G., & Edwards, J. D. (2003). Speed-of-processing and driving simulator training result in improved driving performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*, 218–233.

Rohwedder, S., & Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, *24*, 119–138.

Rosen, A. C., Sugiura, L., Kramer, J. H., Whitfield-Gabrieli, S., & Gabrieli, J. D. (2011). Cognitive training changes hippocampal function in mild cognitive impairment: A pilot study. *Journal of Alzheimer's Disease*, *26*, 349–357.

Rothe, J. P. (1990). *The safety of elderly drivers: Yesterday's young in today's traffic*. New Brunswick, NJ: Transaction Publishers.

Russo, N. M., Hornickel, J., Nicol, T., Zecker, S., & Kraus, N. (2010). Biological changes in auditory function following training in children with autism spectrum disorders. *Behavioral and Brain Functions*, *6*(1), Article 60. doi:10.1186/1744-9081-6-60

Sacks, S., Fisher, M., Garrett, C., Alexander, P., Holland, C., Rose, D., . . . Vinogradov, S. (2013). Combining computerized social cognitive training with neuroplasticity-based auditory training in schizophrenia. *Clinical Schizophrenia & Related Psychoses*, *7*(2), 78–86A.

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*, 203–219.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Salthouse, T. A. (2010). *Major issues in cognitive aging*. Oxford, England: Oxford University Press.

Scalf, P. E., Colcombe, S. J., McCarley, J. S., Erickson, K. I., Alvarado, M., Kim, J. S., . . . Kramer, A. F. (2007). The neural correlates of an expanded functional field of view. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences, 62*(Special Issue 1), 32–44.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Schooler, C., & Mulatu, M. S. (2001). The reciprocal effects of leisure time activities and intellectual functioning in older people: A longitudinal analysis. *Psychology and Aging*, *16*, 466–482.

Schooler, C., Mulatu, M. S., & Oates, G. (1999). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging*, *14*, 483–506.

Schooler, C., Mulatu, M. S., & Oates, G. (2004). Occupational self-direction intellectual functioning, and self-directed orientation in older workers: Findings and implications for individuals and society. *American Journal of Sociology*, *110*, 161–197.

Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoS ONE*, *6*(9), e24372. doi:10.1371/journal.pone.0024372

SharpBrains. (2013). *Web-based, mobile and biometrics-based technology to assess, monitor and enhance cognition and brain functioning: The digital brain health market, 2012–2020*. San Francisco, CA: SharpBrains.

SharpBrains. (2015, January). *The digital brain health market 2012–2020: Web-based, mobile and biometrics-based* *technology to assess, monitor, and enhance cognition and brain functioning* [Addendum]. San Francisco, CA: SharpBrains.

Shatil, E. (2013). Does combined cognitive training and physical activity training enhance cognitive abilities more than either alone? A four-condition randomized controlled trial among healthy older adults. *Frontiers in Aging Neuroscience*, *5*, Article 8. doi:10.3389/fnagi.2013.00008

Shatil, E., Mikulecka, J., Bellotti, F., & Bureš, V. (2014). Novel television-based cognitive training improves working memory and executive function. *PLoS ONE*, *9*(7), e101472. doi:10.1371/journal.pone.0101472

Shimamura, A. P., Berry, J. M., Mangels, J. A., Rusting, C. L., & Jurica, P. J. (1995). Memory and cognitive abilities in university professors: Evidence for successful aging. *Psychological Science*, *6*, 271–277.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. doi:10.1037/a0027473

Siberski, J., Shatil, E., Siberski, C., Eckroth-Bucher, M., French, A., Horton, S., . . . Rouse, P. (2014). Computer-based cognitive training for individuals with intellectual and developmental disabilities: Pilot study. *American Journal of Alzheimer's Disease and Other Dementias*, *30*, 41–48. doi:10.1177/1533317514539376

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Simmons, J. P., Nelson, L., & Simonsohn, U. (2012). A 21 word solution. *Dialogue*, *26*(2), 4–7.

Simons, D. (2013, September 30). Science or sciencey [part 3]. [Blog post]. Retrieved from http://theinvisiblegorilla.com/blog/2010/09/30/science-or-sciencey-part-3/

Simons, D. (2013, September 5). 19 questions about video games, multitasking, and aging (a HI-BAR commentary on a new Nature paper). [Blog post]. Retrieved from http://blog.dansimons.com/2013/09/19-questions-about-video-games.html

Simonton, D. K. (1988). Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin*, *104*, 251–267.

Simonton, D. K. (1990). Creativity and wisdom in aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 320–329). New York, NY: Academic Press.

Simonton, D. K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist*, *55*, 151–158.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, *16*, 97–115.

Singley, M. K., & Anderson, J. R. (1989). *Transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., & Zelinski, E. M. (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.

Smith-Ray, R. L., Hughes, S. L., Prohaska, T. R., Little, D. M., Jurivich, D. A., & Hedeker, D. (2013). Impact of cognitive training on balance and gait in older adults. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences*, *70*, 357–366.

Smith-Ray, R. L., Makowski-Woidan, B., & Hughes, S. L. (2014). A randomized trial to measure the impact of a community-based cognitive training intervention on balance and gait in cognitively intact black older adults. *Health Education & Behavior, 41*(Suppl. 1), 62S–69S.

Sonuga-Barke, E. J., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., . . . European ADHD Guidelines Group. (2013). Nonpharmacological interventions for ADHD: Systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *The American Journal of Psychiatry*, *170*, 275–289.

Spencer-Smith, M., & Klingberg, T. (2015). Benefits of a working memory training program for inattention in daily life: A systematic review and meta-analysis. *PLoS ONE*, *10*(3), e0119522. doi:10.1371/journal.pone.0119522

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., . . . Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, *41*, 638–663.

Stahl, S. M. (1998). Mechanism of action of serotonin selective reuptake inhibitors. Serotonin receptors and pathways mediate therapeutic effects and side effects. *Journal of Affective Disorders*, *51*, 215–235.

Staudinger, U. M. (1996). Wisdom and the social-interactive foundation of the mind. In P. B. Baltes & U. M. Staudinger (Eds.), *Interactive minds* (pp. 276–315). New York, NY: Cambridge University Press.

Staudinger, U. M., & Baltes, P. B. (1996). Interactive minds: A facilitative setting for wisdom-related performance? *Journal of Personality and Social Psychology*, *71*, 746–762.

Staudinger, U. M., Smith, J., & Baltes, P. B. (1992). Wisdom-related knowledge in a life-review task: Age differences and the role of professional specialization. *Psychology and Aging*, *7*, 271–281.

Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology*, *50*, 1049–1059.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*, 2015–2028.

Stevens, C., Fanning, J., Coch, D., Sanders, L., & Neville, H. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Research*, *1205*, 55–69.

Stine-Morrow, E. A. L., & Basak, C. (2011). Cognitive interventions. In K. W. Schaie & S. L. Willis (Eds.), *Handbook of the psychology of aging* (7th ed., pp. 153–170). New York, NY: Elsevier.

Stine-Morrow, E. A. L., Hussey, E. K., & Ng, S. (2015). The potential for literacy to shape lifelong cognitive health. *Policy Insights From the Behavioral and Brain Sciences*, *2*, 92–100. doi:10.1177/2372732215600889

Stine-Morrow, E. A. L., & Payne, B. R. (2015). Education and learning: Lifespan perspectives. In J. D. Wright (Ed.), *International* *encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 7, pp. 137–143). Oxford, England: Elsevier.

Stothart, C. R., Simons, D. J., Boot, W. R., & Kramer, A. F. (2014). Is the effect of aerobic exercise on cognition a placebo effect? *PLoS ONE*, *9*(10): E109557. doi:10.1371/journal.pone.0109557

Strenziok, M., Parasuraman, R., Clarke, E., Cisler, D. S., Thompson, J. C., & Greenwood, P. M. (2014). Neurocognitive enhancement in older adults: Comparison of three cognitive training tasks to test a hypothesis of training transfer in brain connectivity. *NeuroImage*, *85*, 1027–1039.

Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica*, *140*, 13–24.

"Studies on cognitive training benefits." (2014). Retrieved from http://www.cognitivetrainingdata.org/studies-cognitive-training-benefits

Stutts, J. C., Martell, C., & Staplin, L. (2009). *Identifying behaviors and situations associated with increased crash risk for older drivers (DOT HS 811 093)*. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/Traffic%20Injury%20Control/Articles/Associated%20Files/811093.pdf

Subramaniam, K., Luks, T. L., Fisher, M., Simpson, G. V., Nagarajan, S., & Vinogradov, S. (2012). Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. *Neuron*, *73*, 842–853.

Subramaniam, K., Luks, T. L., Garrett, C., Chung, C., Fisher, M., Nagarajan, S., & Vinogradov, S. (2014). Intensive cognitive training in schizophrenia enhances working memory and associated prefrontal cortical efficiency in a manner that drives long-term functional gains. *NeuroImage*, *99*, 281–292.

Swanson, H. L., & Jerman, O. (2007). The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology*, *96*, 249–283.

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*, 2674–2688.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*, 439–471. doi:10.1037/a0033138

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., & Kawashima, R. (2011). Working memory training using mental calculation impacts regional gray matter of the frontal and parietal regions. *PLoS ONE*, *6*(8), e23175. doi:10.1371/journal.pone.0023175

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., . . . Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*, 81–84.

Taya, F., Sun, Y., Babiloni, F., & Beszerianos, A. (2015). Brain enhancement through cognitive training: A new insight from brain connectome. *Frontiers in Systems Neuroscience*, *9*, Article 44. doi:10.3389/fnsys.2015.00044

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., & Gabrieli, J. D. E. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from function MRI. *Proceedings of the National Academy of Sciences, USA*, *100*, 2860–2865.

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall different response time patterns, same retrieval order. *Psychological Science*, *14*, 169–174.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*, 106–113.

Thorndike, E. L. (1906). *The principles of teaching based on psychology*. New York, NY: A. G. Seiler.

Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review*, *21*, 620–628.

"Trained vs. non-trained tasks cont." (2015). Retrieved from http://www.cogmed.com/trained-nontrained-tasks-cont

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*, 446–468.

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is playing video games related to cognitive abilities? *Psychological Science*, *26*, 759–774. doi:10.1177/0956797615570367.

U.S. Food and Drug Administration. (2016, May 19). Step 3: Clinical research. Retrieved from http://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm

Vance, D. E., Dawson, J., Wadley, V., Edwards, J., Roenker, D., Rizzo, M., & Ball, K. (2007). The ACCELERATE study: The longitudinal effect of speed of processing training on cognitive performance of older adults. *Rehabilitation Psychology*, *52*, 89–96.

Vance, D. E., Fazeli, P. L., Ross, L. A., Wadley, V. G., & Ball, K. K. (2012). Speed of processing training with middle-age and older adults with HIV: A pilot study. *Journal of the Association of Nurses in AIDS Care*, *23*, 500–510.

van Ravenzwaaij, D., Boekel, W., Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2014). Action video games do not improve the speed of information processing in simple perceptual tasks. *Journal of Experimental Psychology: General*, *143*, 1794–1805.

Van Vleet, T. M., & DeGutis, J. M. (2013). Cross-training in hemispatial neglect: Auditory sustained attention training ameliorates visual attention deficits. *Cortex*, *49*, 679–690.

Verghese, J., Mahoney, J., Ambrose, A. F., Wang, C., & Holtzer, R. (2010). Effect of cognitive remediation on gait in sedentary seniors. *The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences*, *65*, 1338–1343.

Vinogradov, S., Fisher, M., Holland, C., Shelly, W., Wolkowitz, O., & Mellon, S. H. (2009). Is serum brain-derived neurotrophic factor a biomarker for cognitive enhancement in schizophrenia? *Biological Psychiatry*, *66*, 549–553.

Vinogradov, S., Fisher, M., Warm, H., Holland, C., Kirshner, M. A., & Pollock, B. G. (2009). The cognitive cost of anticholinergic burden: Decreased response to cognitive training in schizophrenia. *The American Journal of Psychiatry*, *166*, 1055–1062.

Von Ah, D., Carpenter, J. S., Saykin, A., Monahan, P., Wu, J., Yu, M., . . . Unverzagt, F. (2012). Advanced cognitive training for breast cancer survivors: A randomized controlled trial. *Breast Cancer Research and Treatment*, *135*, 799–809.

von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, *41*, 611–624.

Voss, M. W., Vivar, C., Kramer, A. F., & van Praag, H. (2013). Bridging animal and human models of exercise-induced brain plasticity. *Trends in Cognitive Sciences*, *17*, 525–544.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*, 129–133. doi:10.1080/00031305.2016.1154108

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*, 470–477.

"Welcome! Let's build your personalized training program." (2013). [Internet archive]. Retrieved from https://web.archive.org/web/20150913171746/https://www.lumosity.com/app/v5/personalization/memory

"Welcome to the brain gym." (2004, February 1). [Blog post]. Retrieved from http://www.israel21c.org/welcome-to-the-brain-gym/

Wertheimer, M. (1959). *Productive thinking*. New York, NY: Harper & Brothers. (Original work published 1945)

Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory—A single-subject analysis. *Physiology & Behavior*, *92*, 186–192. doi:10.1016/j.physbeh.2007.05.041

What Works Clearinghouse. (2010). WWC intervention report: Fast ForWord. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_fastfw_083110.pdf

Willis, S. L. (2001). Methodological issues in behavioral intervention research with the elderly. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 78–108). New York, NY: Academic Press.

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., . . . Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association*, *296*, 2805–2814.

Wilson, R. S., Bennett, D. A., Bienias, J. L., De Leon, C. M., Morris, M. C., & Evans, D. A. (2003). Cognitive activity and cognitive decline in a biracial community population. *Neurology*, *61*, 812–816.

Wolinsky, F. D., Vander Weg, M. W., Howren, M. B., Jones, M. P., & Dotson, M. M. (2013). A randomized controlled trial of cognitive training using a visual speed of processing intervention in middle aged and older adults. *PLoS ONE*, *8*(5), e61624. doi:10.1371/journal.pone.0061624

Wolinsky, F. D., Vander Weg, M. W., Howren, M. B., Jones, M. P., Martin, R., Luger, T. M., . . . Dotson, M. M. (2011). Interim analyses from a randomised controlled trial to improve visual processing speed in older adults: The Iowa Healthy and Active Minds Study. *BMJ Open*, *1*(2), e000225. doi:10.1136/bmjopen-2011-000225

Wu, S., Cheng, C. K., Feng, J., D'Angelo, L., Alain, C., & Spence, I. (2012). Playing a first-person shooter video game induces neuroplastic change. *Journal of Cognitive Neuroscience*, *24*, 1286–1293.

Wu, S., & Spence, I. (2013). Playing shooter and driving videogames improves top-down guidance in visual search. *Attention, Perception, & Psychophysics*, *75*, 673–686.

Zelinski, E. M., Peters, K. D., Hindin, S., & Petway, K. T. (2014). Evaluating the relationship between change in performance on training tasks and on untrained outcomes. *Frontiers in Human Neuroscience*, *8*, Article 617. doi:10.3389/fnhum.2014.00617

Zelinski, E. M., Spina, L. M., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., & Smith, G. E. (2011). Improvement in memory with plasticity-based adaptive cognitive training: Results of the 3-month follow-up. *Journal of the American Geriatrics Society*, *59*, 258–265.

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, *50*, 304–315. doi:10.1037/a0032982